Optimal Dynamic Taxes*

Mikhail Golosov
Yale
and NES

Maxim Troshkin
University of Minnesota
and Federal Reserve Bank of Minneapolis

Aleh Tsyvinski
Yale
and NES

February 2010

Preliminary and incomplete. Do not circulate without permission.

Abstract

This paper provides a methodology to derive simple formulas that facilitate interpretation of the forces behind the optimal taxation results in dynamic settings. The formulas easily connect to empirically observable data and extend the analysis of Diamond (1998) and Saez (2001) to the *dynamic* settings. First, we derive easily interpretable formulas that summarize first-order conditions for dynamic labor and savings distortions in both i.i.d. and persistent shocks cases. Second, we provide numerical simulations of the optimal labor and savings distortions using empirical income distributions and realistic elasticity parameters. Our third contribution is a novel implementation. We show that a tax system based on consolidated income accounts (CIA) implements the optimum. In a given period, the labor income tax depends on labor income and on the balance on the CIA; the savings tax depends only on the amount of savings; the CIA balance is updated as a function of the labor income and the previous balance.

*We thank V.V. Chari, Dirk Krueger, Larry Jones, Stephen Morris, James Poterba, Emmanuel Saez, Nancy Qian, Hongda Xiao, and Pierre Yared for helpful comments

1 Introduction

A sizeable New Dynamic Public Finance (NDPF) literature studies optimal taxation in dynamic settings[1]. The models in this literature extend the classic Mirrlees equity-efficiency trade-offs to dynamic settings in which agents' skills change stochastically over time. The policy relevance of this approach is a subject of debate. On one hand, Banks and Diamond (2008) in a chapter on direct taxation in the Mirrlees Review, argue for importance of the Mirrleesian dynamic and static models as a guide for policy.[2] On the other hand, a recent survey by Mankiw, Weinzierl, and Yagan (2009) of policy-relevance of optimal taxation models states that "*Most of the recommendations of dynamic optimal tax theory are recent and complex*" and that "*The theory of optimal taxation has yet to deliver clear guidance on a general system of ... taxation .... Instead, it has supplied more limited recommendations*". One reason is that the analysis of the dynamic taxation models is often primarily theoretical and uses the language more familiar to a macroeconomist than to a public finance economist. Another reason is that optimal tax systems derived in these models are often difficult to interpret and connect to empirical data of interest in policy applications.

This paper provides a methodology to derive simple formulas that facilitate interpretation of the forces behind the optimal taxation results in dynamic settings. The formulas easily connect to empirically observable data. Diamond (1998) and Saez (2001) derive an easily interpretable formula in terms of elasticities and the shape of income distribution for the *static* Mirrlees models. Our paper extends their analysis to the *dynamic* settings. Most of the analysis in the paper is conducted in a two period model in which the preferences are represented by a utility without income effects. This is primarily done for the ease of exposition. We extend the analysis to the case of more general preferences and multiple periods.

The first contribution is that we derive easily interpretable formulas that summarize first-order conditions for the dynamic labor and savings distortions. Our analysis is

[1] See, for example, Golosov, Kocherlakota, and Tsyvinski (2003) or reviews in Golosov, Tsyvinski, and Werning (2006) and Kocherlakota (2010).

[2] Commissioned by the Institute for Fiscal Studies, the Review is the successor to the influential Meade Report of 1978 and is the authoritative summary of the current state of tax theory as it relates to policy.

conducted both for the i.i.d. shocks and for the persistent shocks. The formulas for the labor distortions are conceptually similar to those derived in the static model. As in the static case, the shape of the income distribution, the redistributionary objectives of the government, and labor elasticity play important roles in the determination of labor distortions. However, the dynamic model adds significant differences to the analysis of optimal distortions.

Consider first the case of the i.i.d. shocks. There are two key insights in this part of the analysis for the nature of labor distortions in the first period (early in life): (a) the dynamic nature of the incentives represents itself as an additional term in the formula for the optimal distortions changing the weights assigned to agents by the social planner, and (b) this reweighing allows to use dynamic incentives to lower marginal taxes for a fraction of sufficiently skilled agents in the first period. The derivation of the easily interpretable formulas for the labor distortions is novel to the New Dynamic Public Finance literature as it primarily focused on the intertemporal (savings) distortion. We then derive a formula representing the savings distortion. The key economic insight of the analysis here is that a high savings distortion should be applied to the high skilled agents as a way to lower their labor distortion. The intuition is that the effort of the highly skilled agents is very valuable in production and deterring their deviations via a savings tax is particularly important. The derivation of the formula interpreting the savings distortion is new to both the NDPF literature, as it allows to study the economic forces determining it, and to the static public finance literature as there is no savings in those models.

We then proceed to the case of persistent shocks. There are two key insights here in addition to the analysis of the static and the i.i.d. case. The first difference is that the optimal labor distortions formulas now depend on conditional rather than on the unconditional distributions of skills. The second insight is that persistence adds an additional force to the optimal tax problem. An agent misrepresenting his skill early in life has under persistent shocks better information than the planner about the true realization of his shocks in the future. This consideration represents itself in changing the weights in the social welfare function assigned to different agents. The planner redistributes away from the types which are more likely to occur if an agent deviated earlier. We must notice that the analysis of the case of the persistent shocks is conducted under the assumption that the first order approach is valid. We verify that this is indeed the case in numerical simulations.

The second contribution of the paper is that we provide numerical simulations for the optimal labor and savings wedges using empirical income distribution and realistic elasticity parameters. The case of the i.i.d. shocks is similar to the numerical analysis of static models in Diamond (1998) and Saez (2001). As in those papers, the labor distortions are U-shaped. There are two key differences with their results. First, we show that the labor distortion for the early period is smaller than for the later period. The main reason for this is that the dynamic provision of incentives allows to decrease distortions early in agents' lives. This result is related to findings in Ales and Maziero (2007) who numerically solve a version of a life cycle economy with i.i.d. shocks drawn from a discrete, two-type distribution, and find that the labor distortions are lower early in life of the household. The second difference from the static model is that we provide calculations for the savings tax and find it numerically significant and increasing. The numerical simulations for the empirically calibrated persistent shocks adds three important differences. The first is that the consideration of conditional rather than the unconditional empirical distributions of income and skills significantly alters the pattern of wedges compared to the static and i.i.d. case. The second difference is that agents face very different labor distortions conditional on the previous shocks. This is due to the effects of the conditional distribution and also of the planner changing weights to deter earlier deviations. Finally, we find that the labor distortions no longer follow a U-shape pattern found in the i.i.d. and in the static models.

The third contribution is a novel implementation of the optimal allocations – consolidated income accounts (CIA) tax system. In a given period, the labor income tax depends on labor income and on the balance of the CIA, the savings tax depends only on the amount of savings; the CIA balance is updated as a function of the labor income and the previous balance. The CIA balance plays a role of the record-keeping device summarizing the previous labor choices of agents. The savings tax is constructed following Werning (2009) as an envelope of the best possible deviation. We then show that the CIA system takes a particularly simple form if the utility is exponential and the shocks are i.i.d. The tax system consists of a non-linear tax on savings income, a non-linear labor income tax, and a CIA account. In each period, a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, updating the CIA balance follow a simple rule. In each period the increase on agent's CIA balance is determined solely by his labor income in that period. Moreover, if the distribution of the shocks does not change over time, taxes do not depend on age of the agent. We conclude with a numerical simulation of the optimal taxes. In our simulation, labor income taxes are lower than labor income distortions as they incorporate the effects of updating the CIA balance.

There are several papers related to our work. The recursive characterization of the problem, especially in the i.i.d. case, has similarities to the Mirrlees (1986) setup with two consumption goods. The first-order approach for persistent shocks is related to the work of Kapicka (2008) who mainly focuses on the risk-sharing properties of the taste shock model with exponential utility and Pareto shocks. There have been very limited theoretical analysis of the labor taxation in dynamic Mirrlees models. One important exception is Battaglini and Coate (2008) who provide a complete characterization of the optimal program with Markovian agents. While incorporating persistence in abilities, most of their analysis for tractability assumes only two ability types and risk neutral individuals.

Numerical simulations in our paper are most related to Weinzierl (2008). He derives an elasticity-based formula in which he studies advantages of a partial reform, a switch to an age-dependent taxation, in a dynamic Mirrlees setting. Albanesi and Sleet (2006) is a comprehensive numerical and theoretical study of optimal capital and labor taxes in a dynamic economy with i.i.d. shocks. Golosov and Tsyvinski (2006) study a disability insurance model with fully persistent shocks. Golosov, Tsyvinski, and Werning (2006) is a two-period numerical study of the determinants of dynamic optimal taxation in the spirit of Tuomala (1990). However, none of these papers, with the exception of Weinzierl (2008), base their analysis on an elasticity-based formula.

Our implementation is related to the work of Albanesi and Sleet (2006), Kocherlakota (2005), Grochulski and Kocherlakota (2007) and, importantly, Werning (2009). Both

Kocherlakota (2005) and Werning (2009) discuss only capital taxation, and our construction of savings taxation builds directly on Werning (2009). The work by Albanesi and Sleet (2006) is another important predecessor to our decentralization. As in our paper, Albanesi and Sleet (2006) keep track of the summary of individuals' past histories summarized by one variable, which is individual's capital stock in their case. There are two main differences between our implementation and that in Albanesi and Sleet (2006). First, their implementation is applicable only for the i.i.d. shocks. Second, in Albanesi and Sleet (2006) savings of households play two roles: intertemporal smoothing and tracking history of previous labor incomes. This allows household to misrepresent their past histories by changing the amount of savings. To prevent households from doing this, the tax on capital should not only be non-linear, but the degree of non-linearity and the amount of capital taxes should depend on the realization of labor income. In our implementation, we keep track of the history of labor incomes in a separate CIA account and do not condition capital tax on income realization which substantially simplifies the tax system, as illustrated in the next section. The implementation we describe is relatively easy to implement in practice as it shares many existing elements with the current US tax code. It also closely related to some long advocated ideas in public finance, such as income tax averaging, which go back to at least to Vickrey (1939) and Vickrey (1947).

The rest of the paper is organized as follows. Section 2 presents the static benchmark model and reviews its main results. Section 3 studies a dynamic economy with i.i.d. shocks. In Section 4 we develop our main results in a dynamic economy with persistent shocks. Section 5 builds a novel decentralization theory based on a tax system with consolidated income accounts. Section 6 concludes.

2 Static benchmark

In this section, we consider a static Mirrlees optimal taxation economy. We review the results by Diamond (1998) and Saez (2001) who analyze the determinants of the marginal tax rates in such economy. This analysis provides a useful benchmark to which we compare our dynamic economy.

Consider a static, one period economy. There is a unit measure of agents. Each agent has a type (skill) $\theta \in \Theta = [0, \infty)$ drawn from a distribution with the distribution function $F(\theta)$ and density $f(\theta)$. Each consumer has preferences $U(c, l)$ where $c \in R_+$ denotes agent's consumption, and $l \in R_+$ denotes agent's effort. We assume that U is strictly concave in c, strictly convex in l, and twice continuously differentiable. Define allocations of effort as $l : \Theta \to R_+$, of consumption as $c : \Theta \to R_+$, and of output as $y : \Theta \to R_+$. An agent of type θ who supplies l units of effort produces $y = \theta l$ units of the output of a consumption good. The type θ of an agent is private information and is known only by the agent. The output y and consumption c are observable.

The aggregate feasibility in this economy satisfies

$$\int c(\theta)\, dF(\theta) \leq \int y(\theta)\, dF(\theta), \tag{1}$$

where $c(\theta)$ and $y(\theta)$ are consumption and output of and agent of type θ.

We focus on the case in which the utility function is given by[3]:

$$U(c, l) = -\frac{1}{\psi} \exp\left(-\psi\left(c - \frac{1}{\gamma} l^{\gamma}\right)\right) \tag{2}$$

where $\psi > 0$, $\gamma > 0$. Here, $\gamma = 1/(1+\alpha)$, where α is the Frisch elasticity of labor supply. This functional form of the utility function implies three simplifying assumptions: (a) there are no income effects, (b) the Frisch elasticity of labor supply is the same for all types, and (c) constant absolute risk aversion. The first two characteristics simplify the formulas for optimal taxes developed in Sections 3. The third characteristic is useful for clearer insights on decentralization in Section 5. As we discuss in Section 4.4, most of the results we develop carry over to more general preferences.

We assume that the government or a social planner is benevolent and has an ability to fully commit ex-ante to a tax system or a mechanism. The planner maximizes a social welfare function $G : R \to R$, where G is concave.

The method for solving this problem is well known starting from the seminal work of Mirrlees (1971)[4], this problem can be solved using a two step procedure. The first step is to represent the problem as a mechanism design problem. The planner receives reports $\sigma(\theta) : \Theta \to \Theta$ from the agents about their types and allocates consumption

[3] Diamond (1998) assumes a similar form of the utility function.

[4] For a textbook treatment see Salanie (2003).

and output $\{c(\theta), y(\theta)\}_{\theta\in\Theta}$ as a function of these reports. The incentive compatibility constraint ensures that no agent would find it in his interest to misrepresent his type:

$$U(c(\theta), y(\theta)/\theta) \geq U(c(\theta'), y(\theta')/\theta) \text{ for all } \theta, \theta'. \tag{3}$$

The constrained efficient (optimal) allocations solve:

$$\max_{\{c(\theta),y(\theta)\}_{\theta\in\Theta}} \int G(U(c(\theta), y(\theta)/\theta))\, dF(\theta) \tag{4}$$

subject to the feasibility constraint (1) and the incentive compatibility constraints (3). Let $\{c^*(\theta), y^*(\theta)\}_{\theta\in\Theta}$ denote the solution to the optimal program (4).

The second step of the solution is *implementation* that characterizes the optimal income taxes $T(y(\theta))$. In the static problem, finding taxes that implement the optimal allocation is straightforward. In particular, the optimal marginal income tax on type θ, $T'(\theta)$, is given by the wedge in the consumption-labor margin:

$$1 - T'(\theta) = \frac{-U_l(c^*(\theta), y^*(\theta)/\theta)}{\theta U_c(c^*(\theta), y^*(\theta)/\theta)},$$

where U_c and U_l denote the derivative of the utility function with respect to c and l.

Diamond (1998) and Saez (2001) show that when preferences satisfy (2), the expression for the optimal marginal taxes can be written as:

$$\frac{T'(\theta)}{1-T'(\theta)} = \gamma \frac{1-F(\theta)}{\theta f(\theta)} \int_\theta^\infty \left(1 - \frac{G'(U(x))\,U(x)}{\lambda}\right) \frac{f(x)}{1-F(\theta)} dx, \tag{5}$$

where

$$\lambda = \int_0^\infty G'(U(x))\,U(x)\,dF(x).$$

The expression (2) is the first order condition for the optimal tax problem that can be used to describe the key determinants of the optimal tax rate. Diamond (1998) and Saez (2001) also use empirical data to calculate the optimal tax rates and to analyze the driving forces of the optimal taxes using this formula.

Expression (5) shows that the optimal marginal taxes in the static economy depend on three key terms. The first term in the expression (5) is related to the elasticity of labor supply. When labor supply is more elastic, marginal taxes are more distortionary. This tends to reduce the magnitude of the optimal marginal tax rates. The second term in the expression(5) , $\frac{1-F(\theta)}{\theta f(\theta)}$, is the ratio between the measure, $1-F(\theta)$, of agents who are more productive than the agent of skill θ and the density at θ, $f(\theta)$. The intuition for the effects of this term on the optimal tax rate is as follows. A positive marginal tax on a type θ is needed to prevent all types above θ to claim the allocation for that type. If the fraction of types more productive than θ is large, a marginal tax on type θ provides stronger incentives to reveal the true type. That calls for higher marginal taxes. On the other hand, if the mass of agents of type θ is large (i.e., $f(\theta)$ is large), or they are sufficiently productive (i.e., θ is large) then the marginal tax on type θ is particularly distortionary, and the planner would choose smaller marginal taxes. Finally, the third term in the expression (5) depends on the curvature in the social welfare function G that captures the degree of redistribution. More concave G tends to raise the third term. Therefore, more redistributionary social preference generally call for higher marginal taxes.

It is important to point out that expression (5) is not a complete closed form solution for the optimal marginal taxes. Rather it is a useful representation of the first order conditions of the optimal problem that provides intuition for the forces affecting taxes. The reason is that the integral on right hand side of expression (5) depends on the optimal level of utility U. For example, consider an increase in γ. Such change has a direct effect on the optimal marginal tax rate by increasing the first term in (5). It also has an indirect effect on the term $G'(U)U$ which formula (5) does not characterize. In practice, formulas like (5) proved to be useful in applications as the insights they provide often closely match the numerical calculations of the optimal taxes (Diamond (1998), Saez (2001), Weinzierl (2008), Golosov, Tsyvinski, and Weinzierl (2009)). Moreover, this formula can be used to prove some more concrete results. We state these results for the ease of a comparison with the dynamic economy. For the details of the proof we refer to Diamond (1998).

Proposition 1. *i). If there exists a skill $\hat{\theta}$ such that skills are distributed with a Pareto distribution with a coefficient a above $\hat{\theta}$, then the optimal marginal taxes $T'(\theta)$ are increasing for all $\theta \geq \hat{\theta}$.*

ii). If, in addition to the conditions in i), the social planner is utilitarian then the *asymptotic marginal taxes satisfy:*

$$\lim_{\theta \to \infty} \frac{T'(\theta)}{1 - T'(\theta)} = \gamma / a. \tag{6}$$

The asymptotic marginal taxes are the same for any social welfare function for which it can be shown that

$$\lim_{\theta \to \infty} G'(U(\theta))U(\theta) = 0$$

iii). Let θ_C be defined from $G'(U(\theta_C))U(\theta_C) = \lambda$. If $\theta f(\theta)$ is increasing on some interval (θ', θ'') for $\theta' \geq \theta_C$ then $T'(\theta)$ is decreasing on (θ', θ'').

The results in the proposition above can be used to show that in a static model calibrated to empirical cross sectional distribution of labor income and tax rates, optimal taxes are of a U-shaped pattern (Diamond (1998), Saez (2001)). That is, the taxes decrease above a certain threshold and then increase. The asymptotic tax rates in such a model are given by (6).

For the analysis that follows it is important to note that the optimal allocations $\{c^*(\theta), y^*(\theta)\}_{\theta \in \Theta}$ can be computed as a solution to the following problem which is dual to the problem (4):

$$V_S(w) = \min_{\{c(\theta), y(\theta)\}} \int (c(\theta) - y(\theta)) \, dF(\theta) \tag{7}$$

subject to the incentive compatibility constraint (3) and

$$w = \int_0^\infty G(U(c(\theta), y(\theta/\theta))) \, dF(\theta), \tag{8}$$

where w is the maximized value of the problem (4), $\int G(U(c^*(\theta), y^*(\theta)/\theta)) \, dF(\theta)$.

For the exposition below it is useful to treat the promised utility w in minimization problem (7) as an arbitrary constant. We now present a useful result that simplifies some of the analysis in the dynamic economy.

Proposition 2. *Let $\{c^*(\theta|w), y^*(\theta|w)\}_{\theta \in \Theta}$ be a solution to (7) for a given w. Suppose the utility function U satisfies (2), and the social welfare function G is homogeneous of degree α. Then the optimal marginal labor distortion $T'(\theta|w)$ defined by*

$$1 - T'(\theta|w) = \frac{-U_l(c^*(\theta|w), y^*(\theta|w)/\theta)}{\theta U_c(c^*(\theta|w), y^*(\theta|w)/\theta)}$$

is independent of w *and for all* $w < 0$*, and*

$$V_S(w) = a_0 - \frac{1}{\alpha\psi}\ln(-w),$$

where a_0 *is a constant.*

Proof. In the appendix. □

3 Optimal allocations in a dynamic economy with i.i.d. shocks

In this section, we extend the analysis of optimal labor taxes developed in the previous section to characterize the optimal labor and savings distortions (implicit taxes) in a dynamic economy. We focus on a two period setting. While many of the arguments extend to arbitrary finite economies, the two-period setting gives particularly stark insights of the characterization of the optimal allocations. This section focuses on the case in which skills in the second period are i.i.d. We first show how to represent the dynamic problem in a form resembling a static Mirrlees problem with two goods. Secondly, we derive an expression similar to (5) for the labor distortions in both periods. There are two key insights in this part of the analysis for the nature of labor distortions in the first period (early in life): (a) the dynamic nature of the incentives represents itself as an additional term in the formula for the optimal distortions changing the weights assigned to agents by the social planner, and (b) this reweighing allows to use dynamic incentives to lower marginal taxes for a fraction of sufficiently skilled agents in the first period. Thirdly, we derive a novel formula similar to (5) that allows to characterize the forces affecting the savings distortions. The key economic insight of the analysis here is that a high savings distortion should be applied to the high skilled agents as a way to lower their labor distortion. The intuition is that the effort of the highly skilled agents is very valuable in production and deterring their deviations via a savings tax is particularly important.

3.1 Analytic characterization

Consider an economy that lasts for two periods: $t = 1, 2$. An agent lives for two periods and has preferences represented by a utility function

$$\mathbb{E}_0 \sum_{t=1,2} \beta^{t-1} U(c_t, l_t),$$

where $c_t \in R_+$ is agent's consumption in period t, $l_t \in R_+$ is agent's labor in period t, $\beta \in (0,1)$ is agent's discount factor, and $\mathbb{E}_0$ is an expectation operator. The instantaneous utility function $U(c_t, l_t)$ is the same as in the benchmark static economy above.

In each period t, agents draw their productivity types, $\theta_t \in \Theta$, from the distribution $F(\theta)$. The draws are independent across periods. Let $\theta^1 = \theta_1$, $\theta^2 = (\theta_1, \theta_2)$ be the history of the shocks. The skill shocks and the history of shocks are privately observed by the agent. Output $y_t = \theta_t l_t$ and consumption c_t are observed by the planner. Let $\Theta^1 = \Theta$ be the set of possible histories in period $t = 1$, and $\Theta^2 = \Theta \times \Theta$ be the set of possible histories in period $t = 2$. Denote by $c_t(\theta^t) : \Theta^t \to \mathbb{R}_+$ agent's allocation of consumption and by $y_t(\theta^t) : \Theta^t \to \mathbb{R}_+$ agent's allocation of output in period t. Denote by $\sigma_t(\theta^t) : \mathbb{R}_+^t \to \mathbb{R}_+$ agent's report in period t. Resources can be transferred between periods with a rate on savings $\delta > 0$. We assume that all savings are publicly observable. Hence, without loss of generality, the social planner controls and chooses the aggregate savings. We also assume that the social planner is utilitarian, and a social welfare function satisfies $G(x) = x$, for all x.

The optimal allocations solve the dynamic mechanism design problem (see, e.g., Golosov, Kocherlakota, and Tsyvinski (2003)):

$$\max_{\left\{c_t(\theta^t), y_t(\theta^t)\right\}_{\theta_t \in \Theta; t=1,2}} \mathbb{E}_0 \left\{U\left(c_1(\theta^1), y_1(\theta^1)/\theta_1\right) + \beta U\left(c_2(\theta^2), y_2(\theta^2)/\theta_2\right)\right\} \tag{9}$$

subject to the incentive compatibility constraint

$$\begin{aligned}
& \mathbb{E}_0 \left\{U\left(c_1(\theta^1), y_1(\theta^1)/\theta_1\right) + \beta U\left(c_2(\theta^2), y_2(\theta^2)/\theta_2\right)\right\} \\
\geq \; & \mathbb{E}_0 \left\{U\left(c_1(\sigma_1(\theta^1)), y_1(\sigma_1(\theta^1))/\theta_1\right) + \beta U\left(c_2(\sigma_2(\theta^2)), y_2(\sigma_2(\theta^2))/\theta_2\right)\right\} \text{ for all } \sigma_t(\theta^t)
\end{aligned}$$

and the feasibility constraint

$$\mathbb{E}_0 \left\{c_1(\theta^1) + \delta c_2(\theta^2)\right\} \leq \mathbb{E}_0 \left\{y_1(\theta^1) + \delta y_2(\theta^2)\right\}.$$

The expectation $\mathbb{E}_0$ above is taken over all possible realizations of histories. The first constraint above is a dynamic incentive compatibility constraint that states that an agent prefers to truthfully report its history of shocks rather than to choose a different reporting strategy. The second constraint is the dynamic feasibility constraint.

Since skills of the agents are i.i.d. across periods, we follow Atkeson and Lucas Jr (1992) and Albanesi and Sleet (2006) to re-write this problem recursively. Here, we briefly describe the recursive formulation and refer to these two paper for the technical details.

Denote by $w(\theta) : \Theta \rightarrow R$ a *promised utility*. In period $t = 1$, the social planner receives agents' reports $\sigma_1(\theta^1)$ and allocates consumption, output, and promised utility $\left\{c\left(\sigma_1\left(\theta^1\right)\right), y\left(\sigma_1\left(\theta^1\right)\right), w\left(\sigma_1\left(\theta^1\right)\right)\right\}_{\theta^1 \in \Theta^1}$. That is, $w\left(\sigma_1\left(\theta^1\right)\right)$ is the expected utility in period $t = 2$ for an agent who sends report $\sigma_1\left(\theta^1\right)$ in period $t = 1$.

The incentive compatibility constraint is given by:

$$U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right) + \beta w\left(\theta\right) \geq U\left(c\left(\theta'\right), y\left(\theta'\right)/\theta\right) + \beta w\left(\theta'\right) \text{ for all } \theta, \theta'. \tag{10}$$

That is, the agent prefers to reveal its true type θ, receive utility $U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right)$ in period $t = 1$ and a continuation utility $w(\theta)$ rather than claim a different type θ'.

In period $t = 1$, the optimal allocations solve the cost minimization problem:

$$\min_{\{c(\theta), y(\theta), w(\theta)\}} \int \left(c\left(\theta\right) - y\left(\theta\right) + \delta V\left(w\left(\theta\right)\right)\right) dF\left(\theta\right) \tag{11}$$

subject to the incentive compatibility constraint (10) and to the promise keeping constraint:

$$w_0 = \int \left(U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right) + \beta w\left(\theta\right)\right) dF\left(\theta\right), \tag{12}$$

where w_0 is a constant. Function $V : R \rightarrow R$ determines the resource cost of delivering a promised utility $w(\theta)$ and is determined by the maximization problem for period $t = 2$.

In period $t = 2$, given the promised utility w, the social planner receives agents' reports $\sigma_2(\theta|w)$ and allocates consumption and output $\left\{c\left(\sigma_2\left(\theta|w\right)\right), y\left(\sigma_2\left(\theta\right)|w\right)\right\}_{\theta \in \Theta}$. The optimal allocations $\left\{c\left(\theta|w\right), y\left(\theta|w\right)\right\}_{\theta \in \Theta}$ for an agent with the assigned promised utility w is a solution to:

$$V\left(w\right) \equiv \min_{\{c(\theta|w), y(\theta|w)\}_{\theta \in \Theta}} \int \left(c\left(\theta|w\right) - y\left(\theta|w\right)\right) dF\left(\theta\right) \tag{13}$$

subject to the incentive compatibility constraint

$$U\left(c\left(\theta|w\right), y\left(\theta|w\right)/\theta\right) \geq U\left(c\left(\theta'|w\right), y\left(\theta'|w\right)/\theta\right) \text{ for all } \theta, \theta'$$

and the promise keeping constraint

$$w = \int U\left(c\left(\theta|w\right), y\left(\theta|w\right)/\theta\right) dF\left(\theta\right)$$

The main goal in this section is to characterize the labor distortion, or wedge in the consumption-labor margin, defined by:

$$1 - T'_{D,t}\left(\theta^t\right) \equiv \frac{-U_l\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)}{\theta_t U_c\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)}. \tag{14}$$

In the benchmark static economy in Section 2, the labor distortion coincides with the marginal income tax rate implementing this distortion. As we discuss in Section 5, this equivalence does not hold in the dynamic setting. For this reason we first characterizing the labor distortions our dynamic economy and then show how to implement such distortions with optimal taxes in Section 5.

In period $t = 2$, it is easy to see that the optimal problem (13) is identical to the static minimization problem (7) (if the objective function of the social planner is given by $G(U) = U$ for all U). The next proposition characterizes the labor distortions in period $t = 2$.

Proposition 3. *Suppose $U(c, l)$ satisfies (2). Then the optimal labor distortion in period $t = 2$ satisfies (5) and is independent of agent's type in period $t = 1$.*

Proof. Follows from (5) and Proposition 2. □

The analysis of the optimal problem (11) in period $t = 1$ is more complex. However, it is conceptually similar to the characterization of a static Mirrlees model with two goods (see, e.g., Mirrlees (1976), Mirrlees (1986), Diamond and Spinnewijn (2009), Golosov, Tsyvinski, and Weinzierl (2009)). We derive the expression for the optimal labor distortions in period $t = 1$ in three steps.

First, we simplify the incentive compatibility by using the envelope theorem. Rewrite the incentive compatibility constraint (10) as a maximization problem. An agent chooses an optimal report θ to maximize the sum of its utility in period $t=1$, $U\left(c\left(\theta'\right), y\left(\theta'\right)/\theta\right)$, and the discounted promised utility, $\beta w\left(\theta'\right)$:

$$u\left(\theta\right) \equiv \max_{\theta' \in \Theta} U\left(c\left(\theta'\right), y\left(\theta'\right)/\theta\right) + \beta w\left(\theta'\right).$$

We apply the envelope theorem to derive the necessary condition for incentive compatibility:

$$u'\left(\theta\right) = -U_l\left(c\left(\theta\right), y\left(\theta\right)/\theta\right)\frac{l\left(\theta\right)}{\theta}. \tag{15}$$

In the Appendix we verify that this condition is also sufficient. When $U\left(c,l\right)$ satisfies (2), (15) becomes:

$$u'\left(\theta\right) = -\psi U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right)\frac{l\left(\theta\right)^{\gamma}}{\theta}. \tag{16}$$

The second step is to re-write the optimal problem (11) such that $u\left(\theta\right)$ is a state variable of optimization. For an agent with the skill θ, utility of reporting the true type is given by $U\left(\theta\right) = U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right)$. Since $u\left(\theta\right) = U\left(\theta\right) + \beta w\left(\theta\right)$, then

$$u'\left(\theta\right) = -\psi\left[u\left(\theta\right) - \beta w\left(\theta\right)\right]\frac{l\left(\theta\right)^{\gamma}}{\theta}.$$

From (2) it follows that:

$$c\left(\theta\right) = \frac{1}{\gamma}l\left(\theta\right)^{\gamma} - \frac{1}{\psi}\ln\left(-\psi U\left(\theta\right)\right). \tag{17}$$

Substituting (16) and (17) in the period $t=1$ optimal program (11), we obtain:

$$\min_{\{l(\theta),u(\theta),w(\theta)\}_{\theta\in\Theta}} \int \left(\frac{1}{\gamma}l\left(\theta\right)^{\gamma} - \frac{1}{\psi}\ln\left(-\psi\left[u\left(\theta\right) - \beta w\left(\theta\right)\right]\right) - \theta l\left(\theta\right) + \delta V\left(w\left(\theta\right)\right)\right) dF\left(\theta\right) \tag{18}$$

subject to

$$u'\left(\theta\right) = -\psi\left[u\left(\theta\right) - \beta w\left(\theta\right)\right]\frac{l\left(\theta\right)^{\gamma}}{\theta}, \tag{19}$$

and

$$w_0 = \int u\left(\theta\right) dF\left(\theta\right).$$

Note that we changed the variables over which we perform minimization. Variables $l\left(\theta\right)$ and $w\left(\theta\right)$ are now control variables, and $u\left(\theta\right)$ is a state variable.

The third step of the analysis is to use standard calculus of variations techniques to characterize (18). We leave the details to the appendix and summarize the main result on the optimal labor distortion in period $t=1$ in the proposition below.

Proposition 4. *Suppose that $U(c,l)$ satisfies (2). Then the optimal labor distortion in period $t=1$ satisfies*

$$\frac{T'_{D,1}(\theta)}{1-T'_{D,1}(\theta)}=\gamma\frac{1-\tilde{F}(\theta)}{\theta\tilde{f}(\theta)}\int_{\theta}^{\infty}\left(1-\frac{U(x)}{\lambda}\right)\frac{\tilde{f}(x)\,dx}{1-\tilde{F}(\theta)} \tag{20}$$

where

$$\tilde{f}(\theta)=\frac{\Psi(\theta)f(\theta)}{\int_0^{\infty}\Psi(x')f(x')\,dx'}, \tag{21}$$

$\Psi(\theta)=\exp\left(\beta\int_0^{\theta}\frac{w'(x)}{U(x)}dx\right)$, $\tilde{F}(\theta)=\int_0^{\theta}\tilde{f}(x)\,dx$, *and*

$$\lambda=\int_0^{\infty}U(x)\,d\tilde{F}(x). \tag{22}$$

Proof. In the Appendix □

Let us compare the optimal labor distortion in period $t=1$ given by (20) to the static labor distortion given by (5). The key difference is in the dynamic provision of incentives. The density of types $f(\theta)$ and the distribution $F(\theta)$ are scaled by a term $\Psi(\theta)$ that depends on the derivative of promised utility, $w'(x)$. Equation (20) shows that the optimal labor distortion in period $t=1$ is the same as the optimal labor distortion in a static economy in which types are drawn from a distribution $\tilde{F}(\theta)$ with the density $\tilde{f}(\theta)$. We can immediately see that for the lowest type in the distribution $\Psi(0)=1$ and that $\Psi'(\theta)=\beta\frac{w'(\theta)}{U(\theta)}\Psi(\theta)$.

In general, it is difficult to determine the sign of marginal promised utility, $w'(\theta)$. It is instructive, however, to consider a case when w is increasing for all θ, and $w'(\theta)>0$. This is indeed the case in all our numerical simulations in Section 3.2. In this case $\Psi'(\theta)\leq 0$ for all θ and distribution has a property:

$$\frac{1-\tilde{F}(\theta)}{\theta\tilde{f}(\theta)}\leq\frac{1-F(\theta)}{\theta f(\theta)},$$

with a strict inequality for interior θ. As we discussed in Section 2, such effect of the distribution adds a force that tends to make marginal taxes lower. Comparing the results in Propositions 3 and 4, we conclude that in a dynamic economy with i.i.d. types there is a force that drives marginal labor distortions lower.

The intuition for this result is as follows. In the static model, marginal labor distortions are used to provide incentives so that the higher types not to choose the allocations of the lower types. In a dynamic economy, there is an additional (dynamic) instrument, promised utility $w(\theta)$, that can be used for the provision of the incentives. Thus, the government can provide incentives with lower labor distortions. This result is related to findings in Ales and Maziero (2007). They numerically solved a version of a life cycle economy with i.i.d. shocks drawn from a discrete, two-type distribution, and found that the labor distortions are lower early in life of the household.

While the sign of $w'(\theta)$ in general is difficult to determine, we can get additional limiting results. The proposition below shows that for sufficiently high skills θ, the marginal labor distortion converges to 0. Since we showed that in the period $t = 2$ the labor distortions are positive and increasing (for example, as is a case for empirically realistic Pareto distributions of skills), the proposition below states that a fraction of agents faces lower marginal taxes in period $t = 1$ compared to their taxes in period $t = 2$.

Proposition 5. *Suppose that $U(c, l)$ satisfies (2). Then $T'_{D,1}(\theta) \to 0$.*

Proof. In the Appendix. □

So far we have focused on the labor distortions in dynamic economies. Another key feature of the optimal allocations in the dynamic Mirrleesian economies typically feature an intertemporal distortion or an implicit savings tax (see, e.g. Golosov, Kocherlakota, and Tsyvinski (2003)). Define the savings wedge as

$$\tau_S(\theta) \equiv 1 - \frac{\delta U_c(c_1(\theta), y_1(\theta)/\theta)}{\beta \mathbb{E}\left\{ U_c\left(c_2(\theta^2), y_2(\theta^2)/\theta^2\right) \Big| \theta \right\}}. \tag{23}$$

The proposition that follows provides a characterization of the savings distortion.

Proposition 6. *Suppose that $U(c, l)$ satisfies (2). Then the savings distortion is positive and satisfies*

$$\tau_S(\theta) = \theta^{\gamma/(\gamma-1)} \frac{\psi}{\gamma} T'_{D,1}(\theta) \left(1 - T'_{D,1}(\theta)\right)^{1/(\gamma-1)}. \tag{24}$$

Proof. In the Appendix □

To understand the implications of this proposition, it is useful to consider the case when γ is large. Then, the last term in (24) is small. The savings distortion is proportional to the labor wedge $T'_{D,1}(\theta)$ and to the skill θ. The intuition is that the labor distortions in period $t = 1$ are the highest for the types that face the most severe incentive problems. Since savings distortions are an additional instrument used to alleviate the incentive problem, it is optimal to have higher distortions on these agents who face the most severe incentive constraints. That is why, savings distortions fall disproportionately on high types. The reason for that is that the same *rate* of labor distortion leads to a larger output loss when applied to high types than to the low types. Therefore it is optimal to substitute from labor distortions to savings distortions when the social planner provides incentives to the more productive types.

Proposition 7. *Suppose that $U(c,l)$ satisfies (2). Then $T'_{D,1}(\theta)$ as well as $U(\theta)$, $w(\theta)$ and $T'_{D,1}(\theta)\, y(\theta)$ all converge to 0. Moreover, if $\frac{\partial}{\partial\theta} T'_{D,1}(\theta)\, y(\theta)$ is bounded from above, $w(\theta)$ is monotonically increasing for all θ sufficiently high.*

3.2 Numerical simulations with i.i.d. shocks

In this section we provide numerical simulations for the optimal labor and savings wedges using micro data under the assumption of i.i.d. shocks. This section parallels the analysis of static models in Diamond (1998) and Saez (2001). As in those papers, our optimal labor distortions appear to be U-shaped. Our analysis differs form theirs in two respects. First, we show that the labor distortion in period $t = 1$ is lower than in period $t = 2$. The main reason is that the dynamic provision of incentives allows decreasing distortions early in agents' lives. The second difference is that we provide calculations for the savings tax and find it numerically significant and increasing.

The data set we use in this section is the Panel Study of Income Dynamics (PSID), 1997 wave.[5] We restrict the sample to the head of household with the total labor income of at least \$10,000. We separate the agents in 10 bins by total labor income: less than \$20,000, from \$20,000 to \$39,999, and so on. In the last bin, we grouped all agents with income above \$180,000.

[5] See http://psidonline.isr.umich.edu/

As in Saez (2001), given the actual tax rates, we next determine the skill distribution generating the labor income of the agents in the sample. We use the data from CBO (2007) and CBO (2008) to obtain effective marginal federal tax rates. The CBO uses the IRS tax returns to construct the effective tax rates. Since the CBO reports effective tax rates for income quintiles, we impute effective tax rates for each of the bins by assuming that the distribution of income in our sample is the same as in the IRS tax returns. That is, we assume the marginal tax rates are equal across all bins falling in a given income quintile corresponding to the CBO data. Then we compute the implied skill for each type (bin) as follows:

$$\theta_i = \frac{Y_i}{\left(Y_i\left(1 - T'\left(Y_i\right)\right)\right)^{1/\gamma}}. \tag{25}$$

In this formula, Y_i is the mean income for each bin (ten thousand dollars). We set $Y_1 = \$15,000$ and $Y_{10} = \$210,000$. Note that with the preferences of the form (2) there are no income effects. The implied skills then can be determined from the static consumption-labor margin in (25). We plot the implied distribution of skills in Figure 1.

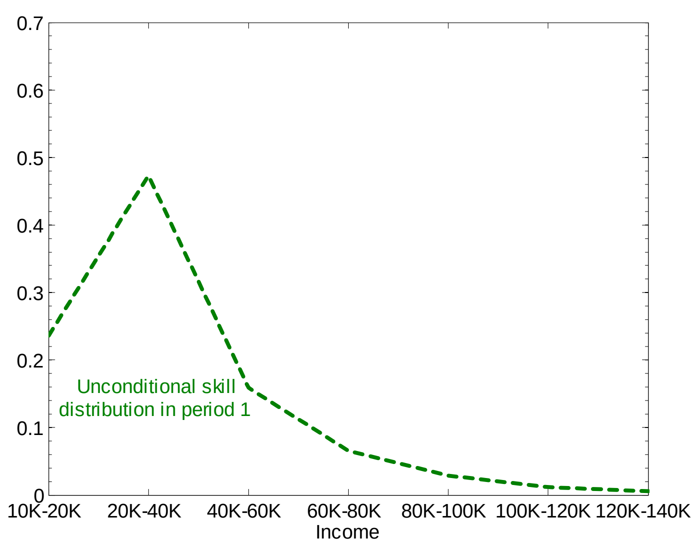

Figure 1: Implied distribution of skills.

We choose $\gamma = 3$, which corresponds to the Frisch elasticity of labor supply equal to 0.5. The coefficient of absolute risk aversion, ψ, is set equal to 10. We set the discount factor $\beta = 0.9852$. We chose the marginal rate of transformation across periods $\delta = 1.015$ so that the social planner at the solution of the optimal program chooses not to transfer resources between the two periods.

To compute the optimal distortions, we first numerically solve for the optimal allocation. We formulate the planner's problem in its primal form (9) and solve it by direct optimization. Our optimization is based on an implementation of the interior-point algorithm that uses conjugate gradient iteration to compute the optimization step. Conjugate gradient iteration offers a way of dealing with possible Jacobian and Hessian singularities. Interior-point approach is one of the most efficient and stable methods that are currently available for solving large nonlinear optimization problems. However, it provides only an approximate estimate of the solution and the optimal set of active constraints. To compute accurate estimates of the solution, including Lagrange multipliers, we switch to an active-set iteration that uses the output of the interior-point algorithm as its input. The implementation of the active-set algorithm that we use is based on the sequential linear quadratic programming. Once we obtain the optimal allocation, the distortions are then computed from their definitions in equations (14) and (23).

The results of the numerical simulation are presented in Figure 2. We report only the values for the first seven bins. There are few observations in the last three bins and thus the implied distribution of skills becomes noisy. Nevertheless, we use all bins for computing the optimal program as the mass of agents and the level of skills in the tail affects the optimal taxes at all skill levels.

First, consider the labor distortions in period $t = 2$ (dash-dotted line in Figure 2). Recall from Proposition 3 that the marginal distortions period $t = 2$ of our dynamic economy with i.i.d. shocks coincide with those in the static economy. The pattern of labor taxes is similar to the results in Diamond (1998) and Saez (2001), who study the optimal taxes in a static Mirrlees economy. The difference is that to find the distribution of labor income Diamond (1998) and Saez (2001) use tax returns data rather than, as we do, the PSID data. Nevertheless, taxes exhibit a U-shaped pattern. The marginal labor distortion for the lowest income (\$10,000-\$19,999) is 27.9 percent. Then the distortion declines and reaches a minimum at 15.9 percent at the second income bin (\$20,000-\$39,999). For the rest of the income bins labor distortion monotonically increases and

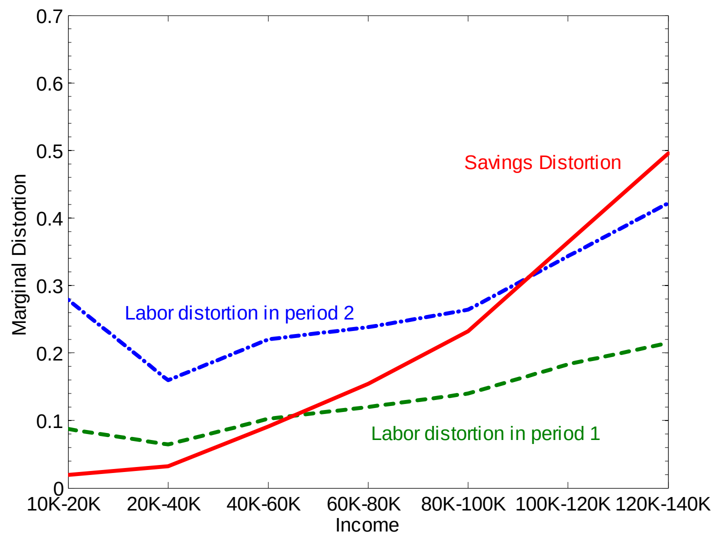

Figure 2: Optimal marginal distortions with i.i.d. shocks.

reaches 42.2 percent for the incomes of \$120,000-\$139,999.

Next, consider the labor distortion in period $t = 1$ (dashed line in Figure 2). We also observe a U-shaped tax, although this pattern is much less pronounced. The marginal labor distortion for the lowest income (\$10,000-\$19,999) is 8.8 percent. Then the distortion declines and reaches a minimum of 6.5 percent at the second income bin (\$20,000-\$39,999). For the rest of the income bins the labor distortion monotonically increases and reaches 21.5 percent for the incomes of \$120,000-\$139,999. An important difference with the static case is that the level of distortions is substantially lower for all income groups, by as much as 10-20 percentage points. As the discussion after Proposition 4 indicates, dynamic provision of incentives enables the planner to lower distortions in period $t = 1$.

Finally, the savings distortion is represented by the solid line in Figure 2. The savings wedge increases for all income levels and ranges from 2 percent for income of \$10,000-\$19,999 to 49.6 percent for income of \$120,000-\$139,999. Notice that the slope increases at incomes of \$20,000-\$39,999. This pattern is consistent with the discussion of the equation (24) for the optimal savings distortion. Above the income of \$20,000-\$39,999 the optimal labor distortion in period $t = 1$ (dashed line) increases as does the level of skills. Equation (24) tells us that both of those effects lead to an increasing savings wedge.

4 Persistent shocks

In this section we extend the analysis of Section 3 to persistent shocks. We show that there are two key insights here in addition to the analysis of the static and the i.i.d. case. The first difference is that the optimal labor distortions formulas now depend on conditional rather than on the unconditional distributions of skills. The second insight is that persistence adds an additional force to the optimal tax problem. An agent misrepresenting his skill early in life has under persistent shocks better information than the planner about the true realization of his shocks in the future. This consideration represents itself in changing the weights in the social welfare function assigned to different agents. The planner redistributes away from the types which are more likely to occur if an agent deviated earlier.

The main difficulty in the analysis is that with persistent shocks it is much harder to find sufficient conditions to ensure that only local incentive constraints bind. We proceed in three steps. First we study a simplified version of the two period economy considered before. An agent can be of only two types in period $t = 1$, but receives a continuum of shocks in period $t = 2$. We set up the recursive formulation of the program and describe the key insights about the effect of the persistent shocks on the optimal labor distortions. Then we extend the analysis to the two period model with continuously distributed shocks in both periods, assuming that only local incentive constraints bind. In the appendix, we derive the sufficient conditions that ensures validity of this approach.

4.1 A model with two shocks

In this subsection, we characterize the labor distortions in a two period economy in which the support of the period $t = 1$ shocks consists of two types. In period $t = 1$, an agent can be one of the two types: θ_H or θ_L, with $\theta_H > \theta_L$. That is, the set of possible histories in period $t = 1$ $\Theta^1 = \{\theta_L, \theta_H\}$.The probability of the realization of each shocks is $f_{1,H}$ and $f_{1,L} = 1 - f_{1,H}$. In period $t = 2$, each of the first period types θ_i for $i \in \{H, L\}$

has a skill shock drawn from the distribution $F_i(\theta)$, where $\theta \in \Theta = [0, \infty)$. Let $f_i(\theta)$ be density function of $F_i(\theta)$. We assume that if $f_H(\theta) > 0$ then $f_L(\theta) > 0$.

The optimal allocations solve the same problem (45), where now the expectation operator $\mathbb{E}_0$ is taken over the distributions described in the previous paragraph. We now proceed to re-write the problem recursively. A recursive formulation of this problem is more involved than in Section 3 as the promised utility is not sufficient to optimally provide incentives. Consider an agent with a high type θ_h in period $t = 1$. If he reports a low shock θ_L to the planner, he knows that his shocks in the second period are drawn from a distribution $F_H(\theta)$ rather than a distribution $F_L(\theta)$. Suppose that, as in the i.i.d. case, the planner assigns the same promised utility $w(\theta_L)$ to the low types who report truthfully and to high types who pretend to be low. The problem is that the shocks in $t = 2$ for these two types agents are drawn from different distributions. For the recursive formulation, we follow the approach developed by Fernandes and Phelan (2000). We now need to assign an additional state variable $w(\theta_L|\theta_H)$. This variable is interpreted as the promised utility for type θ_H who claims to be type θ_L in period 1. [6] The incentive compatibility constraint in $t = 1$ is given by:

$$U(c(\theta_H), y(\theta_H)/\theta_H) + \beta w(\theta_H) \geq U(c(\theta_L), y(\theta_L)/\theta_H) + \beta w(\theta_L|\theta_H). \tag{26}$$

The optimal program for the planner in period $t = 1$ is a modified version of the program for the i.i.d. case (11):

$$\min_{\{c(\theta_i), y(\theta_i), w(\theta_i)\}_{i \in \{L,H\}}, w(\theta_L|\theta_H)} f_{1,H}(c(\theta_H) - y(\theta_H) + \delta V_H(w(\theta_H)))$$
$$+ f_{1,L}(c(\theta_L) - y(\theta_L) + \delta V_L(w(\theta_L), w(\theta_L|\theta_H)))$$

subject to the incentive compatibility constraint (26) and

$$w_0 = \sum_{i \in \{L,H\}} (U(c(\theta_i), y(\theta_i)/\theta_i) + \beta w(\theta_i)) f_{1,i},$$

where $V_H : R \to R$ is a cost of delivering in period $t = 2$ utility $w(\theta_H)$ for the agent who was a high type θ_H in period $t = 1$; $V_H : R^2 \to R$ is a cost of of delivering in period

[6] A general formulation of this problem would also require a promised utility $w(\theta_H|\theta_L)$ for the low types who report that they are high types. Under our interpretation of shocks, such deviations are detectable by the planner and therefore are not binding.

$t=2$ utility $w(\theta_L)$ for the agent who was a low type θ_H in period $t=1$, given that the high type agent who claimed to be a low type in period $t=1$ gets utility $w(\theta_L|\theta_H)$ in period $t=2$. In what follows, to simplify exposition we use a shorthands $w_H, w_L, w_{L|H}$ instead of $w(\theta_H)$, $w(\theta_L)$, and $w(\theta_L|\theta_H)$.

Consider period $t=2$. If an agent reported θ_H in period $t=1$, his allocations solve the maximization problem equivalent to (13) in i.i.d. case. Specifically, the allocations of consumption $c(\theta|w_H)$ and output $y(\theta|w_H)$ solve

$$V_H(w_H) \equiv \min_{\{c(\theta|w_H), y(\theta|w_H)\}_{\theta\in\Theta}} \int (c(\theta|w_H) - y(\theta|w_H))\, dF_H(\theta) \tag{27}$$

subject to the incentive compatibility constraint

$$U(c(\theta|w_H), y(\theta|w_H)/\theta) \geq U(c(\theta'|w_H), y(\theta'|w_H)/\theta) \text{ for all } \theta, \theta',$$

and the promised utility constraint

$$w_H = \int U(c(\theta|w_H), y(\theta|w_H)/\theta)\, dF_H(\theta).$$

Objective functions in the persistent case, equation (27), and the i.i.d. case, equation (13), differ only in the distribution of shocks. In the case of persistent shocks, a conditional probability distribution F_H is used. In the i.i.d. case, the second period shocks are drawn from the distribution F. It is immediate that the optimal labor distortions for the high types in period $t=2$ satisfy (5), where F_H replaces F. The conditional and unconditional distributions may be very different. Hence, the shape of the optimal taxes can differ significantly from the findings based on static Mirrlees models. This difference in distributions is important not just for the theoretical analysis here but also plays a significant role in the numerical simulation in Section 4.3.

In period $t=2$ planner's problem for the types who reported θ_L in the first period is significantly different from (27). The planner promised utility w_L for such type. However, in choosing the allocations $\left\{c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)\right\}_{\theta\in\Theta}$ the social planner also needs to ensure that a type θ_H who misrepresented his type and claimed θ_L in period 1, does not get utility higher than $w_{L|H}$. Therefore, the optimization problem for the types who claimed θ_L is given by:

$$V_L\left(w_L, w_{L|H}\right) \equiv \min_{\left\{c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)\right\}} \int \left(c\left(\theta|w_L, w_{L|H}\right) - y\left(\theta|w_L, w_{L|H}\right)\right) dF_L(\theta) \tag{28}$$

subject to the incentive compatibility constraint

$$U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) \geq U\left(c\left(\theta'|w_L, w_{L|H}\right), y\left(\theta'|w_L, w_{L|H}\right)/\theta\right) \text{ for all } \theta, \theta', \tag{29}$$

the promise keeping constraint

$$w_L = \int U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) dF_L(\theta), \tag{30}$$

and the threat keeping constraint

$$w_{L|H} \geq \int U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) dF_H(\theta). \tag{31}$$

We leave a formal proof of characterization to the Appendix and provide a sketch of the argument. Let p_L and $p_{L|H}$ be Lagrange multipliers on the promise keeping constraint (30) and on the threat keeping constraint (31). Then the optimal program (28) can be re-written as

$$\begin{aligned}
&\min_{\left\{c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)\right\}_{\theta\in\Theta}} \int \left(c\left(\theta|w_L, w_{L|H}\right) - y\left(\theta|w_L, w_{L|H}\right)\right) dF_L(\theta) \\
&+ p_L \int U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) dF_L(\theta) \\
&- p_{L|H} \int U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) dF_H(\theta)
\end{aligned} \tag{32}$$

subject to (29).

We can rearrange (32)

$$\begin{aligned}
&\min_{\left\{c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)\right\}} \int \left(c\left(\theta|w_L, w_{L|H}\right) - y\left(\theta|w_L, w_{L|H}\right)\right) dF_L(\theta) \\
&+ p_L \int \left(1 - \frac{p_{L|H} f_H(\theta)}{p_L f_L(\theta)}\right) U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right) dF_L(\theta)
\end{aligned}$$

Note that in this problem the utility of the agents $U\left(c\left(\theta|w_L, w_{L|H}\right), y\left(\theta|w_L, w_{L|H}\right)/\theta\right)$ is multiplied by the term $\left(1 - \frac{p_{L|H} f_H(\theta)}{p_L f_L(\theta)}\right)$. Intuitively, we can think about this re-written problem as maximizing the pseudo-objective of the planner subject to the incentive constraint of the agent who was type θ_L in period $t = 1$ and reported that type. This pseudo-objective is equivalent to the objective function of a social planner that has (non-normalized) weights $\left(1 - \frac{p_{L|H} f_H(\theta)}{p_L f_L(\theta)}\right)$ instead of the utilitarian weight equal to 1 for all types θ in period $t = 2$. Further inspection shows that the new weights are increasing in the relative probability of receiving a shock θ in period $t = 2$, $f_L(\theta)/f_H(\theta)$. The planner needs to provide utility w_L to the types who were of the true type θ_L (promise keeping) in period $t = 1$. The planner also needs to ensure that the θ_H type who deviated to the low type θ_L does not get utility above $w_{L|H}$ (threat keeping). The efficient way to ensure the threat keeping constraint is to redistribute more to the types in period $t = 2$ which are more likely to occur if agent was indeed low in period $t = 1$.

To see it more clearly, consider a particular process for persistent shocks under which $f_L(\theta)/f_H(\theta)$ is monotonically decreasing in θ. In other words, the higher second period shocks θ are less likely if the agent was θ_L in period $t = 1$. The weight $\left(1 - \frac{p_{L|H} f_H(\theta)}{p_L f_L(\theta)}\right)$ in pseudo-objective function then is monotonically decreasing. In other words, the planner wants to redistribute more to the lower types in period $t = 2$. Consider now a deviation in period $t = 1$. Suppose type θ_H claims to be θ_L. Under our assumption on $f_H(\theta)$, this type is relatively more likely to have high shocks θ in period $t = 2$ and relatively less likely to receive low shocks θ. The social planner who is very redistributive in period $t = 2$ and puts higher (pseudo) weights on the low types allocates relatively low utility to this agent. On the other hand, the type θ_L is not significantly affected, since his probability of having high shocks θ is relatively low. This agent benefits from more redistribution as for him the high shocks θ in the second period are less likely.

We summarize this discussion in the following proposition

Proposition 8. *Suppose $U(c,l)$ satisfies (2). The labor distortion in period $t = 2$ on the type with the first period history $\Theta^1 = \theta_H$ satisfies:*

$$\frac{T'_{D,2}(\theta|\theta_H)}{1 - T'_{D,2}(\theta|\theta_H)} = \gamma \frac{1 - F_H(\theta)}{\theta f_H(\theta)} \int_\theta^\infty \left(1 - \frac{U(x)}{\lambda}\right) \frac{f_H(x)}{1 - F_H(\theta)} dx, \tag{33}$$

where

$$\lambda = \int_0^\infty U(x)\, dF(x).$$

The labor distortion in period $t = 2$ on the type with the first period history $\Theta^1 = \theta_L$ satisfies:

$$\frac{T'_{D,2}(\theta|\theta_L)}{1 - T'_{D,2}(\theta|\theta_L)} = \gamma \frac{1 - F_L(\theta)}{\theta f_L(\theta)} \int_\theta^\infty \left(1 - \frac{\left(1 - \frac{p_{L|H} f_H(x)}{p_L f_L(x)}\right) U(x)}{\lambda}\right) \frac{f_L(x)}{1 - F_L(\theta)} dx, \tag{34}$$

where

$$\lambda = \int_0^\infty \left(1 - \frac{p_{L|H} f_H(x)}{p_L f_L(x)}\right) U(x)\, dF_L(x).$$

4.2 Continuum of shocks

In this section we extend the characterization developed in the previous subsection to the case in which the shocks are drawn from a continuous distribution in the period $t = 1$. The main difficulty in the analysis is finding sufficient conditions that ensure that only local incentive constraints bind in the first period. We proceed in this section by assuming that only local constraints bind and obtain characterization of distortions under such assumption. In the appendix, we provide sufficient conditions. For the simulations in Section 4.3 we verify numerically that this assumption holds.

In periods $t = 1$ and $t = 2$ shocks have the support $\Theta = [0, \infty)$. In period $t = 1$, skills are drawn from the distribution $F(\theta)$. Conditionally on the realization of the shock θ in period $t = 1$, shocks θ' in period $t = 2$ are drawn from the distribution $F(\theta'|\theta)$ with density $f(\theta'|\theta)$.

We now rewrite the problem recursively. For report θ in period $t = 1$, a social planner chooses a function $w(\theta|\theta') : \Theta \times \Theta \to R$ for all possible shocks θ' in period $t = 2$. The function $w(\theta|\theta')$ is a promised utility of a type θ' who reports to the social planner that he is type θ in period $t = 1$. This implies that the incentive compatibility constraint in period $t = 1$ is:

$$U(c(\theta), y(\theta)/\theta) + \beta w(\theta|\theta) \geq U(c(\theta'), y(\theta')/\theta) + \beta w(\theta'|\theta) \text{ for all } \theta, \theta'. \tag{35}$$

The planner's problem in period $t = 1$ is given by:

$$\min_{\{c(\theta), y(\theta), \mathbf{w}(\theta)\}_{\theta \in \Theta}} \int (c(\theta) - y(\theta) + \delta V(\mathbf{w}(\theta), \theta))\, dF(\theta) \tag{36}$$

subject to the incentive compatibility constraint (35) and to the promise keeping constraint

$$w_0 = \int (U(c(\theta), y(\theta)/\theta) + \beta w(\theta|\theta))\, dF(\theta), \tag{37}$$

where $\mathbf{w} : \Theta \to R$ denotes the function $w(\theta|\cdot)$.

The characterization below follows the steps taken in the i.i.d. case.

First, we re-write the incentive compatibility constrain (35) as a maximization problem over agent's types:

$$u(\theta) \equiv \max_{\theta'} U(c(\theta'), y(\theta')/\theta) + \beta w(\theta'|\theta).$$

From the envelope theorem the necessary condition for this problem is given by

$$u'(\theta) = -U_l(c(\theta), y(\theta)/\theta)\frac{l(\theta)}{\theta} + \beta w_2(\theta|\theta). \tag{38}$$

We use $w_1(\theta'|\theta)$ and $w_2(\theta'|\theta)$ to denotes the partial derivative of the promised utility function $w(\theta'|\theta)$ with respect to θ' and θ, respectively.

In general, it is difficult to show under what conditions (38) is also sufficient and can replace (35) in minimization problem (36). In practice, it can be verified numerically (as we do in Section 4.3) if only local constraints bind. When (38) is sufficient, the optimal labor distortions in period $t = 1$ can be found using standard optimal control techniques similar to those used to prove Proposition 4.

When (38) is sufficient to replace the incentive compatibility constraint (35), the maximization problem over the future promised utilities significantly simplifies. The planner only needs to choose the promised utility for the agent who tells his true path $w(\theta|\theta)$ and the derivative of the promised utility, $w_2(\theta|\theta)$. The planner's problem (36) is then given by:

$$\min_{\{c(\theta), y(\theta), w(\theta|\theta), w_2(\theta|\theta)\}_{\theta\in\Theta}} \int (c(\theta) - y(\theta) + \delta V(w(\theta|\theta), w_2(\theta|\theta), \theta))\, dF(\theta) \tag{39}$$

subject to (37) and (38).

It is easy now to write a program for the value function $V(w(\theta_1|\theta_1), w_2(\theta_1|\theta_1), \theta_1)$ in period $t = 2$ for an agent whose type in period $t = 1$ was some value θ_1:

$$V(w, w_2, \theta_1) \equiv \min_{\{c(\theta), y(\theta)\}} \int (c(\theta) - y(\theta))\, dF(\theta|\theta_1) \tag{40}$$

subject to the incentive compatibility constraint (3),

$$w = \int U(c(\theta), y(\theta)/\theta)\, dF(\theta|\theta_1) \tag{41}$$

and the threat-keeping constraint

$$\begin{aligned} w_2 &= \frac{\partial}{\partial \theta_1} \int U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right) dF\left(\theta|\theta_1\right) \\ &= \int U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right) \frac{\partial f\left(\theta|\theta_1\right)}{\partial \theta_1} d\theta. \end{aligned} \tag{42}$$

This problem is a direct generalization of the optimal problem with two types and persistent shocks (28) where the threat-keeping constraint (42) is a continuous type version of (31).

The optimal labor wedges in period $t = 2$ can be found using exactly the steps of Proposition 8. We summarize these findings in the following Proposition.

Proposition 9. *Suppose that $U(c, l)$ satisfies (2) and (38) are sufficient in (36). The labor distortion in the first period satisfy (20), where $\tilde{f}(\theta)$ and λ are given by (21) and (22), and*

$$\Psi\left(\theta\right) = \exp\left(\beta \int_0^\theta \frac{w_1\left(x|x\right)}{U\left(x\right)} dx\right).$$

The labor distortion in the second period on the type who was θ_1 in the first period satisfy

$$\frac{T'_{D,2}\left(\theta|\theta_1\right)}{1 - T'_{D,2}\left(\theta|\theta_1\right)} = \gamma \frac{1 - F\left(\theta|\theta_1\right)}{\theta f(\theta|\theta_1)} \int_\theta^\infty \left(1 - \frac{\left(1 - \frac{p_2 \partial f(x|\theta_1)/\partial \theta_1}{p f(x|\theta_1)}\right) U\left(x\right)}{\lambda}\right) \frac{f\left(x|\theta_1\right)}{1 - F\left(\theta|\theta_1\right)} dx \tag{43}$$

where p and p_2 are constants and

$$\lambda = \int_0^\infty \left(1 - \frac{p_2 \partial f\left(x|\theta_1\right)/\partial \theta_1}{p f\left(x|\theta_1\right)}\right) U\left(x\right) dF\left(x\right).$$

4.3 Numerical simulations with persistent shocks

In this section, we extend the results of Section 3.2 to incorporate persistent shocks. We add observations from the 2005 wave of the PSID to the data constructed in Section 3.2. Since the PSID is a panel data set, with our extended data we are able to compute the distribution of income in 2005 (in 1997 dollars) conditional on income realizations in 1997. We followed the procedure described in Section 3.2 to impute the distribution of skills. We plot the implied skill distributions in Figure 3.

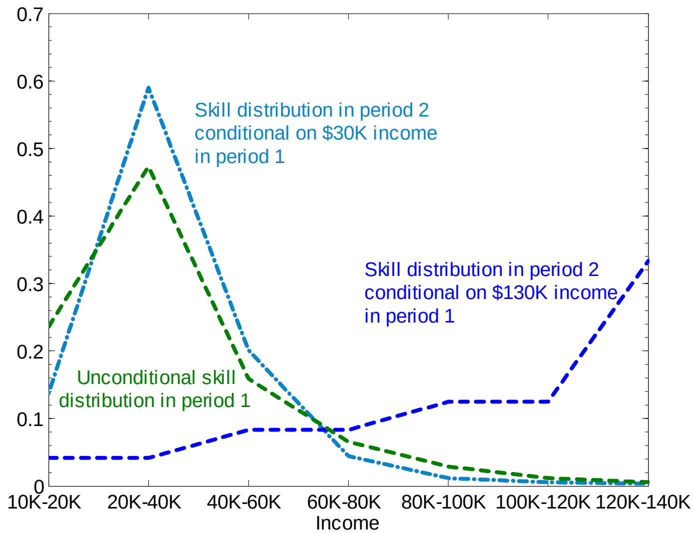

Figure 3: Implied distributions of skills.

Figure 4 displays the results of our numerical simulation. The pattern of labor distortions in period $t = 1$ in the economy with persistent shocks (dashed line in Figure 4) is similar to the case with i.i.d. shocks in Section 3.2. The marginal labor distortion for the lowest income ($10,000-$19,999) is 13.7 percent. The distortion declines and reaches a minimum at 10.3 percent at the second income range ($20,000-$39,999). For the remaining income ranges, the labor distortion monotonically increases and reaches 31 percent for the incomes of $120,000-$139,999. This finding is not surprising. As follows from Proposition 9, the labor distortions in period $t = 1$ are governed by the same expression in both the i.i.d. and persistent shocks cases. The difference comes only in the value of the endogenous variable for the promised utility $w_1(\theta|\theta)$.

Now, consider the optimal labor wedges in period $t = 2$. We plot the wedges for two groups of agents: (a) with income in period $t = 1$ in the $20,000-$39,999 range (lower dot-dashed line in Figure 4), and (b) with income in period $t = 1$ in the $120,000-$139,999 range (higher dot-dashed line in Figure 4).

The first result is that, contrary to the i.i.d. case, these two groups face very different labor distortions in period $t = 2$. The labor distortions of agents who in period $t = 1$ had income of $120,000-$139,999 are much higher (10-40 percent) than their labor distortions

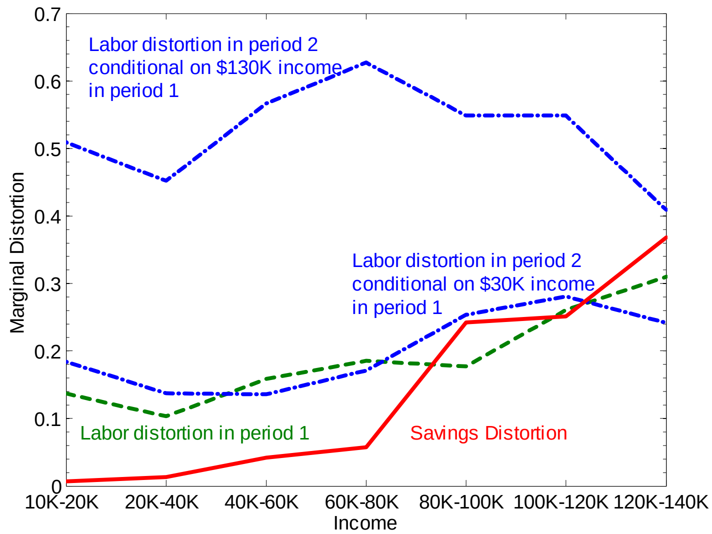

Figure 4: Optimal marginal distortions in the economy with persistent shocks.

in period $t = 1$ (and higher than in the i.i.d. case in Figure 2). The labor distortions for agents who in period $t = 1$ had income of \$20,000-\$39,999 do not change significantly from their earlier distortions (and are lower than in the i.i.d. case). The second result is that the labor distortions no longer follow a U-shape pattern found in the i.i.d. and in the static models. Consider the agents who in period $t = 1$ had income of \$120,000-\$139,999. For such agents, in period $t = 2$ marginal distortions decrease if they have income income above \$60,000.

There are two forces driving the pronounced differences in taxes for these two groups that follow from the discussion of Proposition 8: (a) differences between conditional and unconditional distributions of skills, and (b) differences in conditional distributions across agents, specifically between those who had income of \$120,000-\$139,999 and those who had income of \$20,000-\$39,999.

It follows from (43) that conditional probabilities are some of the key determinants of the optimal marginal labor distortions. Figure 3 displays the conditional probabilities for the two groups of agents we discuss. Consider first an individual in \$20,000-\$39,999 group in period $t = 1$. His income and skill are likely to remain low in the period $t = 2$. The conditional probability of having the same income is just below 60 percent. The conditional distribution for this agent is very concentrated in the first three bins. For those income ranges, the ratio $(1 - F(\theta))/\theta f(\theta)$ of conditional probabilities is lower than for the unconditional probabilities in the i.i.d. case, which is a force driving taxes lower on those types. For higher types the term $1-F(\theta)$ is close to zero, which drives the labor wedge lower than in the case of unconditional distribution even if such individuals make incomes above \$60,000 in period $t = 2$.

Similarly, agents who had income in the \$120,000-\$139,999 range in period $t = 1$ are likely to have high income in period $t = 2$. Therefore, in period $t = 2$, for low θ the term $f(\theta)$ is small while the term $1-F(\theta)$ is large. The ratio $(1 - F(\theta))/\theta f(\theta)$ implies then high marginal taxes. Once agents have high income levels which have substantial conditional probability, labor distortions fall.

The savings distortion is represented by the solid line in Figure 4. The savings wedge increases for all income levels and ranges from less than 1 percent for income of \$10,000-\$19,999 to 36.9 percent for income of \$120,000-\$139,999. For the case of the persistent shocks we do not have an analogue of the equation (24) for the savings wedge that we derived in the i.i.d. case. The overall pattern, however, remains similar with the only difference that the level of the savings distortion is lower.

4.4 Extensions and generalizations

Our analysis so far has focused on a two period economy in which preferences satisfy (2). Although these assumptions simplify the exposition none of them are crucial. It is easy to extend our recursive formulation to the T period model. Then the optimal distortions in any period $t < T$ can be derived using the same steps as we used to derive labor wedges in period $t = 1$ in Propositions 4 and 9.

Similarly, one can extend the analysis to more general utility functions. Because of the exponential utility of the form (2), the benchmark formula for labor distortions in the static case (5) is particularly simple. This formula, however, can be generalized for any preferences that satisfy Spence-Mirrlees single crossing property. For any such preferences, the optimal marginal labor distortion becomes (see Proposition 1 in Saez

(2001)).

$$\frac{T'(\theta)}{1-T'(\theta)} = \left(\frac{1+\gamma^u(\theta)}{\gamma^c(\theta)}\right)\frac{1-F(\theta)}{\theta f(\theta)}$$
$$\times \int_\theta^\infty \exp\left(\int_\theta^x \left(1-\frac{\gamma^u(x')}{\gamma^c(x')}\right)\frac{dx'}{x'}\right)\left(1-\frac{G'(U(x))\,U_c(x)}{\lambda}\right)\frac{f(x)}{1-F(\theta)}dx \tag{44}$$

where $\gamma^u(\theta)$ and $\gamma^c(\theta)$ are uncompensated and compensated elasticities of labor supply of type θ. The same formulas can be derived for the period $t = 2$ of our model for arbitrary preferences (as long as they satisfy the single crossing property). In period $t = 1$, we used exponential preferences to find sufficient conditions for incentives constraints to be replaced by local constraints (Section 7.8). To the extent that such sufficient conditions can be verified, either analytically or numerically, for other preferences, the analysis of the first period wedges and formula (44) can easily be extended to the dynamic case.

5 Consolidated Income Accounts - A Theory of Decentralization

Previous sections characterized the optimal labor and savings distortions (wedges). In dynamic Mirrleesian taxation models, optimal wedges do not necessarily coincide with the taxes implementing the optimal allocations (see, e.g., Golosov and Tsyvinski (2006), Albanesi and Sleet (2006), Kocherlakota (2005)). In this section, we provide a novel implementation of the optimal allocations – a consolidated income accounts (CIA) tax system. In a given period, the labor income tax depends on labor income and on the balance of the CIA, the savings tax depends only on the amount of savings; the CIA balance is updated as a function of the labor income and the previous balance. We then show that the CIA system takes a particularly simple form if the utility is exponential and the shocks are i.i.d. The tax system consists of a non-linear tax on capital income, non-linear labor income tax, and a CIA account. In each period a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, updating the CIA balance follows a simple rule. In each period a change in the CIA balance is determined solely by the individual's labor income in that period. Moreover, if the distribution of the shocks does not change over time, taxes do not depend on age of the agent. We conclude with a numerical simulation of the optimal taxes. In our simulation, labor income taxes are lower than labor income distortions as they incorporate the effects of updating the CIA balance.

5.1 Implementation in a general case

We consider a T period model of the economy described in the previous sections. The optimal allocations $\left\{c^*\left(\theta^t\right), y^*\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}$ solve a generalized version of (9):

$$\max_{\left\{c_t\left(\theta^t\right), y_t\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}} \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c\left(\theta^t\right), y\left(\theta^t\right)/\theta_t\right)\right\} \tag{45}$$

subject to the incentive compatibility constraint

$$\begin{aligned} & \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c\left(\theta^t\right), y\left(\theta^t\right)/\theta_t\right)\right\} \\ \geq \; & \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c\left(\sigma\left(\theta^t\right)\right), y\left(\sigma\left(\theta^t\right)\right)/\theta_t\right)\right\} \text{ for all } \sigma_t\left(\theta^t\right), \end{aligned} \tag{46}$$

and the feasibility constraint

$$\mathbb{E}_0 \sum_{t=1}^{T} \delta^{t-1} c\left(\theta^t\right) \leq \mathbb{E}_0 \sum_{t=1}^{T} \delta^{t-1} y\left(\theta^t\right).$$

As in Kocherlakota (2005), we impose the following assumption on the optimal allocations.

Assumption 1. *For any θ^{t-1}, if $\theta \neq \theta'$ then $y^*\left(\theta^{t-1}, \theta\right) \neq y^*\left(\theta^{t-1}, \theta'\right)$.*

The tax system that we propose consists of three elements. First, taxes depend on a consolidated labor income account (CIA) which keeps track of agents' past earning. We denote by $\omega_t \in R$ the balance on that account, which is updated according to a rule

$$\omega_t = g_t\left(y_t | \omega_{t-1}\right), \tag{47}$$

where $g_t : R_+ \times R \to R$. That is, given a previous balance ω_{t-1} and the current amount of labor income y_t, the agent is assigned a new CIA balance of ω_t according to the function $g_t(\cdot)$. The second element of the tax system is a nonlinear tax on labor income y_t in period t, $T_t(y_t|\omega_t)$, where $T_t : R_+ \times R \to R$. The third element of the tax system is a savings tax $\tau_t(k_t|\omega_t)$, where $\tau_t : R_+ \times R \to R$.

We now formally define the CIA tax system.

Definition 1. *A consolidated income account tax system,* $\{(\omega_t, g_t(y_t|\omega_{t-1})), T_t(y_t|\omega_t), \tau_t(k_t)\}_t$, *consists of,* $\forall t$*: (1) the updating rule for the consolidated income account* $g_t(y_t|\omega_{t-1})$ *and the associated balance on the account* $\omega_t = g_t(y_t|\omega_{t-1})$*, with the initial balance* $\omega_0 = 0$*, (2) the labor income tax* $T_t(y_t|\omega_t)$*, and (3) the savings tax* $\tau_t(k_t)$*.*

We now formally define the notion of implementation.

Definition 2. *A CIA tax system* $\{(\omega_t, g_t(y_t|\omega_{t-1})), T_t(y_t|\omega_t), \tau_t(k_t)\}_t$ implements an allocation $\{\tilde{c}_t(\theta^t), \tilde{y}_t(\theta^t)\}_{\theta^t \in \Theta^t}$ *if this allocation solves*

$$\max_{\{c_t(\theta^t), y_t(\theta^t), k_t(\theta^t)\}_{\theta^t \in \Theta^t}} \mathbb{E}_0 \left\{ \sum_{t=1}^{T} \beta^t U\left(c\left(\theta^t\right), y\left(\theta^t\right)/\theta_t\right) \right\} \tag{48}$$

subject to the budget constraint

$$c_t\left(\theta^t\right) + k_{t+1}\left(\theta^t\right) \le y_t\left(\theta^t\right) + \delta^{-1} k_t\left(\theta^{t-1}\right) - T_t\left(y_t\left(\theta^t\right)|\omega_t\right) - \tau_t\left(k_t\left(\theta^{t-1}\right)\right), \forall t, \ \theta^t \in \Theta^t \text{ and } \theta^t \ge \theta^{t-1}, \tag{49}$$

the updated balances given by (47)*, and* $k_1, \omega_0 = 0$*.*

Note that the agents' problem above assumes that the agent can borrow or lend at the interest rate δ^{-1}. [7] We are now ready to prove the main result of this section that a CIA system implements the optimum. The proof of the theorem shows how to construct this system.

[7] To derive the implications for taxes implementing the optimum one needs to take a stand on what private insurance contract are available. As it is well known in the literature (see, e.g., Prescott and Townsend (1984) or Golosov and Tsyvinski (2006)), if unrestricted multi-period contracts are allowed, the competitive equilibrium allocations are efficient. At the same time, in practice private markets do not provide efficient insurance (see, e.g., Kocherlakota (2005) for some of the arguments). To better understand the implications of the idiosyncratic uncertainty on the optimal non-linear taxes, we consider an important benchmark in which without government intervention no insurance is available to the households beyond borrowing and lending with a risk free rate. This is the most commonly used benchmark in the literature on implementation.

Theorem 1. *Suppose that solution to the optimal program (45), $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$, satisfies Assumption 1. Then there exists a CIA system that implements $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$.*

Proof. We prove this theorem in two steps. The first step is to construct a tax system $\tilde{T}_t\left(y|y_1,...,y_{t-1}\right), \tilde{\tau}_t\left(k\right)$ in which taxes in period t depend on the whole history of labor earnings. We show that such a system implements the solution to (45),$\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$. In the second step, we invoke a standard set theory result to show that the constructed tax system can be equivalently written as $g_t\left(y|\omega_{t-1}\right), T_t\left(y|\omega_{t-1}\right), \tau_t\left(k\right)$. That is, it is sufficient to condition the tax system on the consolidated income account.

Step 1.

We start by recursively constructing the amounts of savings in the optimal allocation:

$$\begin{aligned} k_1^* &= 0, \\ k_t^* &= \delta^{-1}k_{t-1}^* + \mathbb{E}_0\left\{y_t^* - c_t^*\right\},\ \forall t > 1. \end{aligned} \tag{50}$$

For any history $\theta^t \in \Theta^t$, construct the labor tax

$$\tilde{T}_t\left(y_t^*\left(\theta^t\right)|y_1^*\left(\theta^1\right),...,y_{t-1}^*\left(\theta^{t-1}\right)\right) = y_t^*\left(\theta^t\right) - c_t^*\left(\theta^t\right) + \delta^{-1}k_t^* - k_{t+1}^*$$

and set

$$\tilde{T}_t\left(y|y_1^*\left(\theta^1\right),...,y_{t-1}^*\left(\theta^{t-1}\right)\right) = \infty$$

if there exists no $\theta \in \Theta_t$ such that $y = y_t^*\left(\theta^{t-1},\theta\right)$. Note that this tax is well defined because of Assumption 1 since we can invert history of incomes $y_1^*,...,y_{t-1}^*$ into the sequence of shocks θ^{t-1}.

We now construct savings taxes following the contribution by Werning (2009). Let $\tau_t\left(k_t^*\right) = 0$. First, consider the savings tax in period $t = 1$ for savings different from the optimal amount k_1^*. We denote such tax by $\tau_1\left(k_1^* + \varepsilon\right)$ for any $\varepsilon \neq 0$. Let $W_1\left(\varepsilon,\tau\right)$ denote the highest ex-ante utility that any agent can achieve who saves ε additional units of capital in period $t=1$ on which he pays tax τ in period $t=2$:

$$
\begin{aligned}
&W_1(\varepsilon,\tau_\varepsilon) = \\
&\qquad \sup_{\sigma(\theta^T)\in\Sigma} \mathbb{E}_0 \left\{ U\left(c^*\left(\sigma\left(\theta^1\right)\right) - \varepsilon, \frac{y^*\left(\sigma\left(\theta^1\right)\right)}{\theta_1} \right) \right. \\
&\qquad \left. + \beta U\left(c^*\left(\sigma\left(\theta^2\right)\right) + \delta^{-1}\varepsilon - \tau, \frac{y^*\left(\sigma\left(\theta^2\right)\right)}{\theta_2} \right) \right\} \\
&\qquad + \mathbb{E}_0 \left\{ \sum_{t=3}^{T} \beta^{t-1} U\left(c^*\left(\sigma\left(\theta^t\right)\right), \frac{y^*\left(\sigma\left(\theta^t\right)\right)}{\theta_t} \right) \right\}
\end{aligned}
$$

Note from (46) that

$$
W_1(0,0) = \mathbb{E}\left\{ \sum_{t=1}^{T} \beta^{t-1} U\left(c^*\left(\theta^t\right), y^*\left(\theta^t\right)/\theta_t\right) \right\}.
$$

Since $W_1(\varepsilon,\tau)$ is monotonically decreasing in τ, there exists τ_ε^* such that the agent is indifferent between the best deviation (for a given amount of savings in period $t=1$, ε) and the optimal allocation:

$$
W_1(\varepsilon,\tau_\varepsilon^*) = W_1(0,0).
$$

Set $\tau_2(k_1^*+\varepsilon) = \tau_\varepsilon^*$ for all ε.

Similarly, define by $W_2(\varepsilon,\tau)$ the utility under best deviation (for a given amount of savings ε) for period $t=2$:

$$
\begin{aligned}
W_2(\varepsilon,\tau) = &\sup_{\sigma(\theta^T)\in\Sigma,\varepsilon_1} \mathbb{E}_0 \left\{ U\left(c^*\left(\sigma\left(\theta^1\right)\right) - \varepsilon_1, \frac{y^*\left(\sigma\left(\theta^1\right)\right)}{\theta_1} \right) \right. \\
&\left. + \beta U\left(c^*\left(\sigma\left(\theta^2\right)\right) + \delta^{-1}\varepsilon_1 - \tau_2(k_1^*+\varepsilon_1) - \varepsilon, \frac{y^*\left(\sigma\left(\theta^2\right)\right)}{\theta_2} \right) \right\} \\
&+ \mathbb{E}_0 \left\{ \beta^2 U\left(c^*\left(\sigma\left(\theta^3\right)\right) + \delta^{-1}\varepsilon - \tau, \frac{y^*\left(\sigma\left(\theta^3\right)\right)}{\theta_3} \right) \right. \\
&\left. + \sum_{t=4}^{T} \beta^{t-1} U\left(c^*\left(\sigma\left(\theta^t\right)\right), \frac{y^*\left(\sigma\left(\theta^t\right)\right)}{\theta_t} \right) \right\}
\end{aligned}
$$

Given the amount of savings in period $t=2$, ε, and the tax in period $t=3$, τ, the agent chooses the reporting strategy $\sigma\left(\theta^T\right)\in\Sigma$ and the savings in period $t=1$, ε_1. The agent pays the savings tax $\tau_2(k_1^*+\varepsilon_1)$ on savings ε_1.

As before, we find τ_ε^* such that an agent is indifferent between the best deviation (for a given amount of savings in period $t = 2$, ε) and the optimal allocation:

$$W_2(\varepsilon, \tau_\varepsilon^*) = W_1(0,0)$$

We then set the savings tax in period $t = 3$, $\tau_3(k_2^* + \varepsilon) = \tau_\varepsilon^*$. We proceed by induction to define $\tau_t(k_t)$ for all $t > 1$.

Now we verify that constructed tax system $\left\{\tilde{T}_t, \tau_t\right\}_t$ implements the optimum. Consider any sequence of labor incomes $y_t\left(\theta^t\right)$, $\forall \theta^t \in \Theta^t$. We want to show that for any such sequence, and the equilibrium choice of consumption $c\left(\theta^t\right)$ and savings k_t, agent's utility is weakly less than utility from choosing the optimal allocation $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}$ and the corresponding sequence $\{k_t^*\}_t$. First, we can restrict our attentions to the sequences $y_t\left(\theta^t\right)$ that satisfy $y_t\left(\theta^t\right) = y_t^*\left(\hat{\theta}^t\right)$ for some $\hat{\theta}^t$. From the definition of the labor taxes, all other sequences involve labor taxes sufficiently large that an agent does not choose them. Construction of capital taxes implies that the agent's choice of savings satisfies $k_t = k_t^*$, for all t. This implies that the agent's choice of consumption $c_t\left(\theta^t\right) = c_t^*\left(\hat{\theta}^t\right)$. Let $\hat{\sigma}_t\left(\theta^t\right) = \hat{\theta}^t$ for all t. Then the utility of the agent from such a labor choice is

$$\mathbb{E}_0\left\{\sum_{t=1}^{T} \beta^t U\left(c\left(\hat{\sigma}\left(\theta^t\right)\right), y\left(\hat{\sigma}\left(\theta^t\right)\right)/\theta_t\right)\right\}$$

which, by (46), is less then his utility of optimal allocation.

$$\mathbb{E}_0\left\{\sum_{t=1}^{T} \beta^t U\left(c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)/\theta_t\right)\right\}.$$

In other words, the above reasoning is as follows. Labor taxes are constructed such that agent chooses income from the menu of the optimal allocations under different strategies. Capital taxes are constructed such that the agent's savings and consumption are among the menu of the optimal allocations. Finally, the incentive compatibility constraint insures that the best choice of such consumption and labor menu is indeed the optimal allocation.

Step 2.

We now construct a CIA tax system. Recall that a t dimensional space of real numbers is equivalent to $\mathbb{R}$, i.e., there is a bijection $q_t : \mathbb{R}^t \to \mathbb{R}$ (see, e.g. Problem 9 on page 20 in Kolmogorov and Fomin (1975)). Let $\omega_t = q^t(y_1, ..., y_t)$ and denote by q_t^{-1} the inverse of q_t. Define taxes as

$$T_t(y|\omega_{t-1}) = \tilde{T}_t\left(y|q_{t-1}^{-1}(\omega_{t-1})\right)$$

and

$$\tau_t(k|\omega_{t-1}) = \tilde{\tau}_t\left(k|q_{t-1}^{-1}(\omega_{t-1})\right).$$

Finally define a consolidated labor income account as

$$g_t(y|\omega_{t-1}) = q_t\left(q_{t-1}^{-1}(\omega_{t-1}), y\right).$$

It is clear from the construction that this system implements the same allocations as the tax system constructed in Step 1. □

Theorem 1 is quite general. In particularly, as long as Assumption 1 is satisfied, the theorem does not depend on the details of the stochastic process for skills or the form of preferences and holds irrespective of whether only local incentive constraints bind. However, despite the simplicity of construction, the tax functions not necessarily would be well behaved with respect to ω_t. For example, the proof does not say anything about whether such functions are continuous, monotonic, or differentiable.

Theorem 1 establishes that the planner needs relatively limited amount of information to design taxes that implement the optimal allocation. It is possible to set up a rule for the consolidated income account that keeps track of a one dimensional summary statistics of the past labor incomes, together with nonlinear taxes on labor (which is a function of that summary statistics) and savings taxes to achieve the constrained-efficient allocations.

5.2 Implementation in a simple case

At the level of generality in the previous section, it is difficult to characterize the CIA rule g_t or the dependence of labor taxes on the CIA balance $T_t(y_t|\omega_t)$ in more details. We turn to a special case of our model which allows to understand the properties of a CIA system in simple cases.

Proposition 10. *Suppose θ_t are i.i.d. with the distribution $F_t(\theta)$, and the utility satisfies (2). Then there exists a CIA tax system that implements the optimal allocation and satisfies*

$$T_t(y_t|\omega_{t-1}) = T_t(y_t) - \omega_{t-1},$$

and

$$\omega_t = g_t(y_t) + \omega_{t-1}.$$

Moreover, if $F_t(\theta) = F(\theta)$ for all $t < T$ then $g_t = g$ and $T_t(\cdot) = T(\cdot)$ for all $t < T$.

Proof. In the Appendix □

This Proposition shows that when shocks and utility satisfy assumptions of Proposition 10, the tax system takes a particularly simple form. It consists of a non-linear tax on capital income, a non-linear labor income tax, and a CIA account. In each period a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, rules for the updating the CIA balance follow a simple rule. In each period the increase on your account is determined solely by the individual's labor income in that period.

The last part of the proposition also clarifies the relationship between history dependence (as embodied in the CIA account) and the age-dependence taxes. Proposition 10 shows that if the shocks up are drawn from the same distribution, the tax and CIA rules do not need to be conditioned on age explicitly. The exception is the last period of agents' life, since it is no longer possible to provide incentives dynamically. One can think of period T as the retirement period, in which shocks θ_T are being drawn from a "post-retirement" distribution $F_T(\theta)$. In that case Proposition 10 shows that the retirement benefits should be governed by somewhat different rules then "pre-retirement" taxes. One particularly interesting case to consider is the one where all mass point of skill distribution is on zero, $F_T(0) = 1$, so that no agent can work after retirement. In that case, the taxes in period T reduce to $T_T = -\omega_T$. Obviously, T_T are interpreted as the retirement benefits, which are equal to the balance on the CIA accumulated over your lifetime.

We compute T and g functions for the parameters described in Section 3.2 and present them in Figure 5. Marginal labor tax functions T' and marginal CIA function g' are represented by solid lines. To facilitate comparison, optimal marginal labor distortions are also reproduced from Section 3.2. Labor distortion in period $t = 1$ is represented by a dashed line. Labor distortion in period $t = 2$ is represented by a dot-dashed line that coincides with $T_2'(y_2)$.

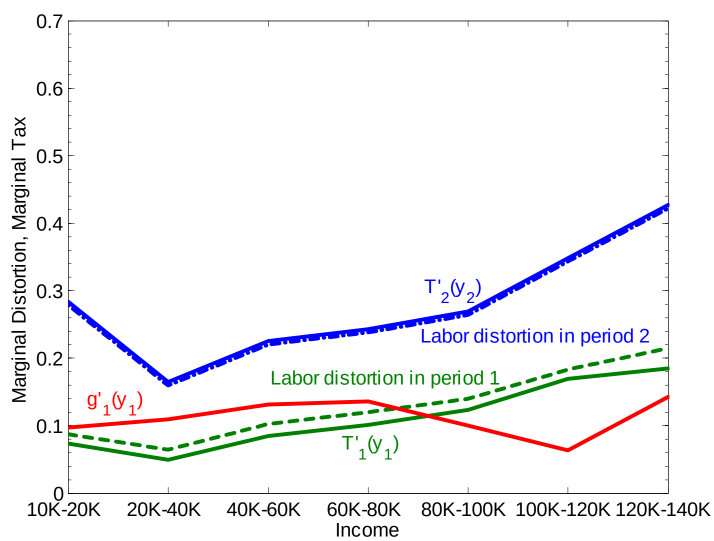

Figure 5: Implementation with a CIA tax system.

First, notice is that the marginal labor taxes in period 1, $T_1'(y)$ are lower than the marginal labor distortions (wedges) $T_{D,1}'(y)$. To understand this result, consider the first order conditions to (48):

$$U_y \frac{1}{\theta} = (1 - T_1'(y_1)) U_c + \zeta g_1'(y_1)$$

where ζ is a Lagrange multiplier on the balance of the CIA (47). As we saw from Figure 2, in the optimal allocation agents with higher y_1 receive higher period $t = 2$ utility, which, from construction of g_1, implies that $g_1'(y_1) > 0$. Therefore

$$T_1'(y_1) < T_{D,1}'(y_1).$$

The labor taxes are lower than then the ones characterized in equation (20). This relationship is not surprising. The labor choice of individuals is distorted by both the marginal tax rate on labor income, and by the distortions arising from the CIA. The CIA is used to provide incentives dynamically – agents with higher income are assigned a higher CIA balance. The sum of the two distortions is equal to the optimal labor distortion characterized in Section 3. Therefore, the marginal tax rate are lower than the optimal labor distortions.

We now compare our decentralization to the ideas of cumulative income tax averaging proposed by Vickrey (1939) and Vickrey (1947). One of the main motivations behind Vickrey's proposal was to avoid *"inequality of burden as between taxpayers of fluctuating and of steady incomes"*. Although the implementation details of our proposal differ from those of Vickrey's, conceptually they are closely related. Table 1 compares two individuals, with a similar present value realization of labor income, and the total labor taxes they pay. The present value of taxes for these individuals are similar, which is in line with Vickrey's motivation for the cumulative average income tax. The details differ. Under Vickrey's proposal, an individual in period $t = 2$ should pay tax based on the lifetime present value of the labor income. In our decentralization, the labor tax in period $t = 2$ is on labor income in that period, but the first individual gets a bigger CIA credit for the higher taxes he paid in period $t = 1$. Our system is a natural extension of Vickrey since it allows to provide better insurance against idiosyncratic shocks by incorporating savings distortions.

Table 1. Two realizations of labor income, labor taxes, and CIA balances (optimal allocation is computed in Section 3.2, units are normalized so that $y_1(\theta_{10}) = 1$)

	θ_1	θ_2	$y_1(\theta_1) + \delta y_2(\theta_2)$	$T_1(y_1) + \delta T_2(y_2)$	$g_1(y_1)$
Realization 1:	10	9	1.66	0.64	0.14
Realization 2:	9	10	1.63	0.58	0.09

In the more general case, when shocks do not follow the i.i.d. process, the characterization of the taxes and CIA balances implementing the optimal allocations is likely to be significantly more complicated. We explore this issues at length in Golosov, Troshkin, and Tsyvinski (2009).

6 Conclusion

This paper provides a methodology to study the determinants of the optimal taxes using the first order conditions of the optimal mechanism design problem. We derive a novel implementation using the consolidated income accounts. We also provide numerical simulations for the simple cases of the optimal taxation problem and the taxes implementing it. An important next step is to explore more general calculations of the optimum and the taxes implementing it. Such calculations for the general case of the multi-period model with persistent shocks are very complex. Instead, a different approach can be used. Following, important contributions of Conesa and Krueger (2006), Conesa, Kitao, and Krueger (2009) one can use parametrized simple tax functions. The advantage of the current paper is that it gives insights on how to design such tax function, what parameters and which functions to include. We use these insights and provide realistic computations for a calibrated model in Golosov, Troshkin, and Tsyvinski (2009).

7 Appendix (very preliminary and incomplete)

7.1 Proof of Proposition 2

First we prove that the value function takes the form $V_S(w) = \ln(-w) + a_0$. Let $a_0 = V_S(-1)$. Since any $w < 0$ can be represented as $-1\exp(\ln(-w))$, constraint (8) becomes

$$\begin{aligned}
-1 &= \int_0^\infty \exp(-\ln(-w))\, G(U(c(\theta), y(\theta/\theta)))\, dF(\theta) \\
&= \int_0^\infty G\left(\exp\left(-\frac{1}{\alpha}\ln(-w)\right) U(c(\theta), y(\theta/\theta))\right) dF(\theta) \\
&= \int_0^\infty G\left(U\left(c(\theta) + \frac{1}{\psi\alpha}\ln(-w), y(\theta/\theta)\right)\right) dF(\theta)
\end{aligned}$$

where in the second line we used the fact the G is homogeneous of degree α and in the third that U satisfies (2). Let $\tilde{c}(\theta) = c(\theta) + \frac{1}{\alpha\psi}\ln(-w)$, so that

$$-1 = \int_0^\infty G(U(\tilde{c}(\theta), y(\theta/\theta)))\, dF(\theta) \tag{51}$$

Since U satisfies (2), (3) is equivalent to

$$U(\tilde{c}(\theta), y(\theta)/\theta) \geq U(\tilde{c}(\theta'), y(\theta')/\theta) \text{ for all } \theta, \theta'. \tag{52}$$

The objective function in (7) can be re-written as

$$V_S(w) = \max_{\{\tilde{c}(\theta), y(\theta)\}} \int (\tilde{c}(\theta) - y(\theta))\, dF(\theta) - \frac{1}{\alpha\psi}\ln(-w)$$

s.t. (51) and (52), so that

$$V_S(w) = a_0 - \frac{1}{\alpha\psi}\ln(-w).$$

Note that this implies that $y^*(\theta|w)$ that solves (7) is independent of w. Since U satisfies (2), the marginal labor distortion satisfies

$$1 - T'(\theta|w) = \frac{(y^*(\theta|w))^{\gamma-1}}{\theta^\gamma}$$

and it is also independent of w.

7.2 Proof of Proposition 4

Rewrite problem (18) as

$$\max_{\{l(\theta),u(\theta),w(\theta)\}} -\int \left(\frac{1}{\gamma}l\left(\theta\right)^{\gamma}-\frac{1}{\psi}\ln\left(-\psi\left[u\left(\theta\right)-\beta w\left(\theta\right)\right]\right)-\theta l\left(\theta\right)+\delta V\left(w\left(\theta\right)\right)\right)dF\left(\theta\right) \tag{53}$$

s.t.

$$u'\left(\theta\right)=-\psi\left[u\left(\theta\right)-\beta w\left(\theta\right)\right]\frac{l\left(\theta\right)^{\gamma}}{\theta}$$

and

$$w_0=\int u\left(\theta\right)dF\left(\theta\right)$$

Set up Hamiltonian

$$\begin{aligned}\mathcal{H} &= -\left(\frac{1}{\gamma}l\left(\theta\right)^{\gamma}-\frac{1}{\psi}\ln\left(-\psi\left[u\left(\theta\right)-\beta w\left(\theta\right)\right]\right)-\theta l\left(\theta\right)+\delta V\left(w\left(\theta\right)\right)\right)f\left(\theta\right)\\ &+ pu\left(\theta\right)f\left(\theta\right)-\mu(\theta)\psi\left[u\left(\theta\right)-\beta w\left(\theta\right)\right]\frac{l\left(\theta\right)^{\gamma}}{\theta}\end{aligned}$$

Recalling that $U\left(\theta\right)=u\left(\theta\right)-\beta w\left(\theta\right)$, the optimality condition for u is

$$\left(\frac{1}{\psi U\left(\theta\right)}+p\right)f\left(\theta\right)-\mu\left(\theta\right)\psi\frac{l\left(\theta\right)^{\gamma}}{\theta}=-\mu'\left(\theta\right)$$

Integrating,

$$\begin{aligned}\mu\left(\theta\right) &= \int_{\theta}^{\infty}\exp\left(-\int_{\theta}^{x}\psi\frac{l\left(\theta'\right)^{\gamma}}{\theta'}d\theta'\right)\left(\frac{1}{\psi U(x)}+p\right)f\left(x\right)dx\\ &= \int_{\theta}^{\infty}\frac{1}{\psi U(x)}\exp\left(-\int_{\theta}^{x}\psi\frac{l\left(\theta'\right)^{\gamma}}{\theta'}d\theta'\right)\left(1+p\psi U\left(x\right)\right)f\left(x\right)dx\end{aligned}$$

This implies that

$$U\left(\theta\right)\mu\left(\theta\right)=\int_{\theta}^{\infty}\frac{1}{\psi}\frac{U\left(\theta\right)}{U(x)}\exp\left(-\int_{\theta}^{x}\psi\frac{l\left(\theta'\right)^{\gamma}}{\theta'}d\theta'\right)\left(1+p\psi U\left(x\right)\right)f\left(x\right)dx \tag{54}$$

Consider the term $\frac{U(\theta)}{U(x)}\exp\left(-\int_{\theta}^{x}\psi l\left(\theta'\right)^{\gamma}/\theta' d\theta'\right)$:

$$\frac{U\left(\theta\right)}{U(x)}\exp\left(-\int_{\theta}^{x}\psi l\left(\theta'\right)^{\gamma}/\theta' d\theta'\right)=\exp\left(\ln\left(\frac{U\left(\theta\right)}{U(x)}\right)-\int_{\theta}^{x}\psi l\left(\theta'\right)^{\gamma}/\theta' d\theta'\right) \tag{55}$$

Since

$$\begin{aligned}\ln U(x) - \ln U(\theta) &= \int_\theta^x \frac{d\ln U(\theta')}{d\theta'} d\theta' \\ &= \int_\theta^x \frac{U'(\theta')}{U(\theta')} d\theta' \\ &= -\int_\theta^x \left(\psi \frac{l(\theta')^\gamma}{\theta'} + \frac{\beta w'(\theta')}{U(\theta')}\right) d\theta'\end{aligned}$$

where in the last line we used (19) and $U'(\theta) = u'(\theta) - \beta w'(\theta)$. Then (55) becomes

$$\begin{aligned}\frac{U(\theta)}{U(x)} \exp\left(-\int_\theta^x \psi l(\theta')^\gamma / \theta' d\theta'\right) &= \exp\left(-\int_\theta^x \frac{d\ln U(\theta')}{d\theta'} d\theta' - \int_\theta^x \psi l(\theta')^\gamma / \theta' d\theta'\right) \\ &= \exp\left(\int_\theta^x \frac{\beta w'(\theta')}{U(\theta')}\right)\end{aligned}$$

Substitute it into (54)

$$U(\theta)\mu(\theta) = \int_\theta^\infty \frac{1}{\psi} \exp\left(\int_\theta^x \frac{\beta w'(\theta')}{U(\theta')}\right)(1 + p\psi U(x)) f(x)\, dx \tag{56}$$

Since the boundary condition is $\mu(0) = 0$, (56) implies that

$$\lambda \equiv -\frac{1}{p\psi} = \frac{\int_0^\infty \exp\left(\int_0^x \frac{\beta w'(\theta')}{U(\theta')}\right) U(x) f(x)\, dx}{\int_0^\infty \exp\left(\int_0^x \frac{\beta w'(\theta')}{U(\theta')}\right) f(x)\, dx} \tag{57}$$

Now consider the optimality conditions for $l(\theta)$

$$-\left(l(\theta)^{\gamma-1} - \theta\right) f(\theta) = \mu(\theta)\psi\left(u(\theta) - \beta w(\theta)\right)\gamma \frac{l(\theta)^{\gamma-1}}{\theta} \tag{58}$$

Since labor distortion $T'_{D,1}(\theta)$ satisfies

$$T'_{D,1}(\theta) = 1 - l^{\gamma-1}/\theta, \tag{59}$$

(58) becomes

$$\frac{T'_{D,1}(\theta)}{1 - T'_{D,1}(\theta)} = \frac{\mu(\theta)\psi U(\theta)\gamma}{\theta f(\theta)} \tag{60}$$

Substitute (57) and (56) into (60) to get

$$\frac{T'_{D,1}(\theta)}{1 - T'_{D,1}(\theta)} = \frac{\gamma}{\theta f(\theta)} \int_\theta^\infty \exp\left(\int_\theta^x \frac{\beta w'(\theta')}{U(\theta')}\right)\left(1 - \frac{U(x)}{\lambda}\right) f(x)\, dx \tag{61}$$

Let $\Psi(x) = \exp\left(\int_0^x \frac{\beta w'(\theta')}{U(\theta')}\right)$ and

$$\tilde{f}(x) = \frac{\Psi(x) f(x)}{\int_0^\infty \Psi(x') f(x') dx'}$$

Function $\tilde{f}(x)$ is a well defined density function, and we denotes its cdf by $\tilde{F}(x)$.

Then (61) becomes

$$\begin{aligned}
\frac{T'_{D,1}(\theta)}{1 - T'_{D,1}(\theta)} &= \frac{\gamma}{\theta f(\theta) \exp\left(\int_0^\theta \frac{\beta w'(\theta')}{U(\theta')}\right)} \int_\theta^\infty \exp\left(\int_0^x \frac{\beta w'(\theta')}{U(\theta')}\right)\left(1 - \frac{U(x)}{\lambda}\right) f(x) dx \\
&= \frac{\gamma}{\theta \tilde{f}(\theta)} \int_\theta^\infty \left(1 - \frac{U(x)}{\lambda}\right) \tilde{f}(x) dx \\
&= \gamma \frac{1 - \tilde{F}(\theta)}{\theta \tilde{f}(\theta)} \int_\theta^\infty \left(1 - \frac{U(x)}{\lambda}\right) \frac{\tilde{f}(x) dx}{1 - \tilde{F}(\theta)}
\end{aligned}$$

where

$$\lambda = \int_0^\infty U(x) d\tilde{F}(x)$$

7.3 Proof of Proposition 6

When utility function satisfies (2), (23) becomes

$$\tau_S(\theta) = 1 - \frac{\delta U(c_1(\theta), y_1(\theta)/\theta)}{\beta \mathbb{E}\left\{U\left(c_2(\theta^2), y_2(\theta^2)/\theta^2\right) \middle| \theta\right\}}.$$

>From the envelope theorem applies to problem (13), it can be shown that

$$\frac{1}{V'(w(\theta))} = -\psi \mathbb{E}\left\{U\left(c_2(\theta^2), y_2(\theta^2)/\theta^2\right) \middle| \theta\right\}.$$

Consider the optimality condition for $w(\theta)$ in problem (53).

$$\left(-\frac{\beta}{\psi U(\theta)} - \delta V'(w(\theta))\right) f(\theta) = -\mu(\theta) \beta \psi \frac{l(\theta)^\gamma}{\theta}$$

or

$$\tau_S(\theta) = 1 - \frac{\delta}{\beta} \psi V'(w(\theta)) U(\theta) = \mu(\theta) U(\theta) \psi^2 \frac{l(\theta)^\gamma}{\theta f(\theta)}.$$

Since $U(\theta)$ is negative, and, from (58), $\mu(\theta)$ is negative, this implies that the savings distortion is positive. Substitute (60)

$$\tau_S(\theta) = \psi \frac{1}{\gamma} \frac{T'_{D,1}(\theta)}{1 - T'_{D,1}(\theta)} l(\theta)^\gamma.$$

Substitute (59) to get

$$\begin{aligned}\tau_S(\theta) &= \theta^{\gamma/(\gamma-1)}\frac{\psi}{\gamma}\frac{T'_{D,1}(\theta)\left(1-T'_{D,1}(\theta)\right)^{\gamma/(\gamma-1)}}{1-T'_{D,1}(\theta)} \\ &= \theta^{\gamma/(\gamma-1)}\frac{\psi}{\gamma}T'_{D,1}(\theta)\left(1-T'_{D,1}(\theta)\right)^{1/(\gamma-1)}\end{aligned}$$

7.4 Proof of Proposition 7

To be added

7.5 Proof of Proposition 8

Equation (33) follows directly from the discussion in the text. Here we prove (34). Set up Hamiltonian for (32) (we omit conditioning on w_L and $w_{H|L}$ for conciseness).

$$\begin{aligned}\mathcal{H} &= -\left(\frac{1}{\gamma}l(\theta)^\gamma - \frac{1}{\psi}\ln\left(-\psi\left[u(\theta)-\beta w(\theta)\right]\right) - \theta l(\theta)\right)f_L(\theta) \\ &+ p_L u(\theta) f_L(\theta) - p_{L|H}u(\theta) f_H(\theta) - \mu(\theta)\psi\left[u(\theta)-\beta w(\theta)\right]\frac{l(\theta)^\gamma}{\theta} \\ &= -\left(\frac{1}{\gamma}l(\theta)^\gamma - \frac{1}{\psi}\ln\left(-\psi\left[u(\theta)-\beta w(\theta)\right]\right) - \theta l(\theta)\right)f_L(\theta) \\ &+ p_L\left(1-\frac{p_{L|H}f_H(\theta)}{p_L f_L(\theta)}\right)u(\theta) f_L(\theta)\end{aligned}$$

The rest of the steps follow those in the proof of Proposition 4.

7.6 Proof of Proposition 9

The derivation of the second period distortions follows exactly the steps of the proof of Proposition 8. To derive the first period labor distortion, we follow the steps of the proof of Proposition 4 and set up Hamiltonian

$$\begin{aligned}\mathcal{H} &= -\left(\frac{1}{\gamma}l(\theta)^\gamma - \frac{1}{\psi}\ln\left(-\psi\left[u(\theta)-\beta w(\theta)\right]\right) - \theta l(\theta) + \delta V\left(w(\theta|\theta), w_2(\theta|\theta), \theta\right)\right)f(\theta) \\ &+ p u(\theta) f(\theta) - \mu(\theta)\left(\psi\left[u(\theta)-\beta w(\theta|\theta)\right]\frac{l(\theta)^\gamma}{\theta} - \beta w_2(\theta_2|\theta_2)\right).\end{aligned}$$

Using optimality condition on $u(\theta)$ and integrating $\mu'(\theta)$ we again obtain (54) and

$$\frac{U(\theta)}{U(x)}\exp\left(-\int_\theta^x \psi l(\theta')^\gamma/\theta' d\theta'\right)=\exp\left(-\int_\theta^x \frac{d\ln U(\theta')}{d\theta'}d\theta'-\int_\theta^x \psi l(\theta')^\gamma/\theta' d\theta'\right). \tag{62}$$

Since $u'(\theta)=U'(\theta)+\beta w_1(\theta|\theta)+\beta w_2(\theta|\theta)$,

$$\begin{aligned}\frac{d\ln U(\theta)}{d\theta}=\frac{U'(\theta)}{U(\theta)}&=\frac{u'(\theta)-\beta w_1(\theta|\theta)-\beta w_2(\theta|\theta)}{U(\theta)}\\&=\frac{-\psi\left[u(\theta)-\beta w(\theta|\theta)\right]\frac{l(\theta)^\gamma}{\theta}+\beta w_2-\beta w_1(\theta|\theta)-\beta w_2(\theta|\theta)}{U(\theta)}\\&=-\psi\frac{l(\theta)^\gamma}{\theta}-\frac{\beta w_1(\theta|\theta)}{U(\theta)}\end{aligned} \tag{63}$$

Substitute (63) into (62) to get

$$\frac{U(\theta)}{U(x)}\exp\left(-\int_\theta^x \psi l(\theta')^\gamma/\theta' d\theta'\right)=\exp\left(\int_\theta^x \frac{\beta w_1(\theta|\theta)}{U(\theta)}\right).$$

The rest of the proof follows the proof of Proposition 4.

7.7 Proof of Proposition 10

First, consider the recursive formulation of the optimal problem

$$V_t(w)=\min_{\{c(\theta),y(\theta),w'(\theta)\}}\int\left(c(\theta)-y(\theta)+\delta V_{t+1}(w'(\theta))\right)dF_t(\theta) \tag{64}$$

s.t. (10) and

$$w=\int\left(U(c(\theta),y(\theta)/\theta)+\beta w'(\theta)\right)dF_t(\theta).$$

By Proposition 2, $V_T(w)=a_T-\frac{1}{\psi}\ln(-w)$ where a_T is a constant. First we want to show that $V_t(w)=a_t-\frac{1}{\psi}\ln(-w)$ for all t where a_t are constants. The proof of this fact is by induction. Suppose that $V_{t+1}(w)=a_{t+1}-\frac{1}{\psi}\ln(-w)$ and let $a_t=V_t(-1)$. Let $\tilde{c}(\theta)=c(\theta)+\frac{1}{\psi}\ln(-w)$ and $\tilde{w}'(\theta)=-w'(\theta)/w$. Substitute the newly defined variables to (65) taking into account that U satisfies (2):

$$V_t(w)=\min_{\{\tilde{c}(\theta),y(\theta),\tilde{w}'(\theta)\}}\int\left(\tilde{c}(\theta)-y(\theta)+\delta V_{t+1}(\tilde{w}'(\theta))\right)dF_t(\theta)+\frac{1}{\psi}\ln(-w) \tag{65}$$

s.t.

$$U(\tilde{c}(\theta),y(\theta)/\theta)+\beta\tilde{w}'(\theta)\geq U(\tilde{c}(\theta'),y(\theta')/\theta)+\beta\tilde{w}'(\theta')$$

and

$$1=\int\left(U(\tilde{c}\left(\theta\right),y\left(\theta\right)/\theta)+\beta\tilde{w}'\left(\theta\right)\right)dF_t\left(\theta\right).$$

Solution to (65) coincides with the solution to $V_t(-1)$.

For the stuff below we note that this implies that for any two $\breve{w},\hat{w}$,

$$\frac{\hat{w}}{\breve{w}}=\frac{w'\left(\theta;\hat{w}\right)}{w'\left(\theta;\breve{w}\right)} \tag{66}$$

Now we define the CIA tax system. First we define all variables as a function of shocks, and then we invert them. Let

$$\tilde{\omega}_t\left(\theta^t\right)=\frac{1}{\psi}\ln\left(-w'\left(\theta^t\right)\right)$$

and have $\tilde{g}_t\left(\tilde{\omega}_{t-1},\theta_t\right)$ satisfy

$$\tilde{\omega}_t\left(\theta^t\right)=\tilde{g}_t\left(\tilde{\omega}_{t-1},\theta_t\right). \tag{67}$$

Consider two histories, $\breve{\theta}^t$ and $\hat{\theta}^t$ with a property that $\breve{\theta}_t=\hat{\theta}_t$. Then

$$\begin{aligned}\tilde{\omega}_t\left(\breve{\theta}^t\right)-\tilde{\omega}_t\left(\hat{\theta}^t\right)&=\frac{1}{\psi}\ln\left(\frac{w'\left(\breve{\theta}^t\right)}{w'\left(\hat{\theta}^t\right)}\right)\\&=\frac{1}{\psi}\ln\left(\frac{w'\left(\breve{\theta}^{t-1}\right)}{w'\left(\hat{\theta}^{t-1}\right)}\right)\\&=\tilde{\omega}_t\left(\breve{\theta}^{t-1}\right)-\tilde{\omega}_t\left(\hat{\theta}^{t-1}\right)\end{aligned}$$

where we used (66) on the second line. This implies that (68) can be re-written

$$\tilde{\omega}_t\left(\theta^t\right)=\tilde{g}_t\left(\theta_t\right)+\tilde{\omega}_{t-1}. \tag{68}$$

Finally, we are ready to define the CIA system. Let

$$\omega_t\left(y\left(\theta^t\right)\right)=\tilde{\omega}_t\left(\theta^t\right)$$

and

$$g_t\left(\omega_{t-1}\left(y\left(\theta^{t-1}\right)\right),y\left(\theta_t\right)\right)=\tilde{g}_t\left(\tilde{\omega}_{t-1}\left(\theta^{t-1}\right),\theta_t\right).$$

>From (68), g_t and ω_t satisfy[8]

$$\omega_t\left(y\right) = g_t\left(y\right) + \omega_{t-1}.$$

Define k_t^* as in (50). Finally, we define taxes on labor as

$$\tilde{T}_t\left(\theta^{t-1},\theta_t\right) = y_t\left(\theta^{t-1},\theta_t\right) - c_t\left(\theta^{t-1},\theta_t\right) + \delta^{-1}k_t^* - k_{t+1}^* - \tilde{\omega}_{t-1}\left(\theta^{t-1}\right) + A_t$$

where A_t is a constant. Since $y_t\left(\theta^{t-1},\theta_t\right)$ does not depend on θ^{t-1} since $y\left(\theta^{t-1},\theta_t\right)$ is a solution to (65), and $c_t\left(\theta^{t-1},\theta_t\right) + \tilde{\omega}_{t-1}\left(\theta^{t-1}\right)$ do not depend on θ^{t-1} by construction, $\tilde{T}_t$ depends only on the current realization of θ_t. Therefore, we can set

$$T_t\left(y\left(\theta_t\right)\right) = \tilde{T}_t\left(\theta^{t-1},\theta_t\right).$$

Finally we choose A_t so that

$$A_t = \int \tilde{\omega}_{t-1}\left(\theta^{t-1}\right) dF\left(\theta^{t-1}\right).$$

7.8 Sufficiency of local incentive constraints

To be added.

[8] If y_t is chosen such that there is no history θ for which $y\left(\theta\right)$ is a solution, we can set $g_t(y) = \underline{\omega}$ where $\underline{\omega}$ is a sufficiently low number

References

Albanesi, S., and C. Sleet (2006): "Dynamic optimal taxation with private information," *Review of Economic Studies*, 73(1), 1–30.

Ales, L., and P. Maziero (2007): "Accounting for private information," *working paper*.

Atkeson, A., and R. Lucas Jr (1992): "On efficient distribution with private information," *The Review of Economic Studies*, 59(3), 427–453.

Banks, J., and P. Diamond (2008): "The Base for Direct Taxation," *Mirrlees Review: Reforming the tax system for the 21st century*.

Battaglini, M., and S. Coate (2008): "Pareto efficient income taxation with stochastic abilities," *Journal of Public Economics*, 92(3-4), 844–868.

CBO (2007): "Historical Effective Federal Tax Rates, 1979 to 2005," *Congressional Budget Office*.

——— (2008): "Historical Effective Federal Tax Rates, 1979 to 2005: Supplement with Additional Data on Sources of Income and High-Income Households," *Congressional Budget Office*.

Conesa, J., S. Kitao, and D. Krueger (2009): "Taxing capital? Not a bad idea after all!," *American Economic Review*, 99(1), 25–48.

Conesa, J., and D. Krueger (2006): "On the optimal progressivity of the income tax code," *Journal of Monetary Economics*, 53(7), 1425–1450.

Diamond, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

Diamond, P., and J. Spinnewijn (2009): "Capital Income Taxes with Heterogeneous Discount Rates," *NBER Working Paper*.

Fernandes, A., and C. Phelan (2000): "A recursive formulation for repeated agency with history dependence," *Journal of Economic Theory*, 91(2), 223–247.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

Golosov, M., M. Troshkin, and A. Tsyvinski (2009): "A Quantitative Exploration in the Theory of Dynamic Optimal Taxation," *working paper*.

Golosov, M., and A. Tsyvinski (2006): "Designing optimal disability insurance: A case for asset testing," *Journal of Political Economy*, 114(2), 257–279.

Golosov, M., A. Tsyvinski, and M. Weinzierl (2009): "Preference Heterogeneity and Optimal Commodity Taxation," *working paper*.

Golosov, M., A. Tsyvinski, and I. Werning (2006): "New dynamic public finance: A user's guide," *NBER Macroeconomics Annual*, 21, 317–363.

Grochulski, B., and N. Kocherlakota (2007): "Nonseparable Preferences and Optimal Social Security Systems," *NBER Working Paper*.

Kapicka, M. (2008): "Efficient allocations in dynamic private information economies with persistent shocks: A first order approach," *working paper*.

Kocherlakota, N. (2005): "Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation," *Econometrica*, 73(5), 1587–1621.

Kocherlakota, N. (2010): *The New Dynamic Public Finance*. Princeton University Press, USA, forthcomming.

Kolmogorov, A., and S. Fomin (1975): *Introductory real analysis*. Dover publications.

Mankiw, N., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *NBER Working Paper*.

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

——— (1976): "Optimal tax theory: A synthesis," *Journal of Public Economics*, 6(4), 327–358.

——— (1986): "The theory of optimal taxation," *Handbook of mathematical economics*, 3, 1197–1249.

Prescott, E., and R. Townsend (1984): "Pareto optima and competitive equilibria with adverse selection and moral hazard," *Econometrica*, 52(1), 21–45.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

Salanie, B. (2003): *The economics of taxation*. The MIT press.

Tuomala, M. (1990): *Optimal income tax and redistribution*. Oxford University Press, USA.

Vickrey, W. (1939): "Averaging of income for income-tax purposes," *The Journal of Political Economy*, pp. 379–397.

Vickrey, W. (1947): *Agenda for progressive taxation*. The Ronald Press Company.

Weinzierl, M. (2008): "The surprising power of age-dependent taxes," *working paper*.

Werning, I. (2009): "Nonlinear Capital Taxation," *working paper*.